
Proceedings of:

Uncertainty Analysis in Measurement and Design

2004 Summer Topical Meeting

June 30 and July 1, 2004

*The Nittany Lion Inn
The Pennsylvania State University
State College, Pennsylvania*

The American Society for Precision Engineering (ASPE) is a multidisciplinary professional and technical society concerned with research and development, design, manufacture and measurement of high accuracy components and systems. ASPE activities encompass relevant aspects of mechanical, electronic, optical and production engineering, physics, chemistry, and computer and materials science. Membership is open to anyone interested in any aspect of precision engineering.

Founded in 1986, ASPE provides a new focus for a diverse but important community. Other professional organizations have covered aspects of precision engineering, always as a sideline to their principal goals. ASPE is based on the core of generic concepts necessary to achieve precision in any application; independent of discipline, ASPE intends to be the focus for precision technology - and to represent all facets from research to application.

*ASPE -- The American Society for Precision Engineering
301 Glenwood Avenue, Suite 205, Raleigh, NC 27603
P.O. Box 10826, Raleigh, NC 27605
Telephone: (919) 839-8444 Fax (919) 839-8039
Web site: <http://www.aspe.net>*

Preface

This book comprises the proceedings of the 2004 Summer Topical Meeting. The contributions reflect the authors' opinions and are published as presented to ASPE, without change. Their inclusion in this publication does not necessarily constitute endorsement by the ASPE, or its editorial staff.

Individual readers of this book and nonprofit libraries acting for them are permitted to make fair use of the material contained in this book to be used in teaching or research; provided they copy said paper in total, as written by the author. Permission is granted to quote excerpts from the scientific or technical works in this book with proper acknowledgment of the source to include: The author's name, the book name, and the ASPE tag line. Reproduction of figures and tables is likewise permitted from the contents of this book, provided that the same acknowledgment of the same source and content is permitted.

Any other type reproduction, systematic copying or multiple reproduction of any material in this book is expressly prohibited except with the written permission of the ASPE editorial staff, addressed to the Society's headquarters.

ISBN 1-887706-34-8

2004

*American Society for Precision Engineering
301 Glenwood Avenue, Suite 205
Raleigh, NC 27603*

Printed in the United States of America

2004 Summer Topical Meeting

Contents

<i>Preface</i>	<i>ii</i>
<i>Contents</i>	<i>iii</i>
<i>Program</i>	<i>iv</i>
<i>Foreword</i>	<i>v</i>
<i>Technical Paper Index</i>	<i>vi</i>
<i>Technical Papers</i>	3-151
<i>Author Index</i>	153

2004 Summer Topical Meeting

Meeting Schedule

	Wednesday, June 30		Thursday, July 1
7:45	<i>Registration & Light Breakfast</i>	7:45	<i>Registration & Light Breakfast</i>
8:30	<i>TECHNICAL SESSION I</i>	8:30	<i>TECHNICAL SESSION V</i>
10:00		10:00	
	<i>Break</i>		<i>Break</i>
10:30	<i>TECHNICAL SESSION II</i>	10:30	<i>TECHNICAL SESSION VI</i>
12:00		12:00	
	<i>Lunch on your own</i>		<i>Lunch on your own</i>
1:00		1:00	
	<i>POSTER SESSION</i>		<i>POSTER SESSION</i>
1:30	<i>TECHNICAL SESSION III</i>	1:30	<i>TECHNICAL SESSION VII</i>
3:00		3:00	
	<i>Break</i>		<i>Wrap-up</i>
3:30	<i>TECHNICAL SESSION IV</i>	3:30	
4:30			
5:30	<i>Tour of Machine Dynamics Research Lab</i>		
6:30			

Foreword

This meeting will focus on the theoretical and practical aspects of uncertainty analysis, motivated by the growing acceptance of the ISO Guide to the Expression of Uncertainty in Measurement. The intent of the meeting is to discuss current research and practice, with a particular focus on the needs of industry. Presentations will emphasize both fundamental research and practical applications. Talks with significant tutorial content are also planned and encouraged.

Uncertainty analysis is a vital tool in measurement, inspection, and precision machine/instrument design. When thoughtfully and properly applied, such analysis allows the metrologist to provide a meaningful statement of measurement quality and the designer to predict machine performance and to minimize cost. The meeting will deal with all aspects of evaluating the uncertainty of measurements and constructing realistic uncertainty budgets in support of the design process.

The objective of the meeting is to bring together researchers and practitioners from industry, government, and academia to provide a forum for the exchange of ideas. In addition to paper and poster presentations, we will have discussions on current industry needs and the perceived relationship between those needs and current research.

It is anticipated that this subject will attract participants who are not normally active in the ASPE. We encourage the attendance of non-ASPE members and look forward to their contributions.

2004 Summer Topical Meeting Co-Chairmen:

W. Tyler Estler

National Institute of Standards and Technology

Eric R. Marsh

The Pennsylvania State University

Technical Sessions

Session I

Standards Activities in Measurement Uncertainty

Wednesday, June 30, 8:30 AM – 10:00 AM

1. **The Joint Committee for Guides in Metrology: Supplementing the Guide to the Expression of Uncertainty in Measurement** page 3
Estler, W. T. (National Institute of Standards and Technology)
2. **Measurement Uncertainty and Traceability Issues: A Standards Activity Update** page 5
Phillips, S. D. (National Institute of Standards and Technology)
3. **Uncertainty Due to Finite Resolution Measurements** Page 14
Phillips, S. D.; Tolman, B.; Estler, W. T. (National Institute of Standards and Technology)

Session II

Uncertainty Issues in Industry

Wednesday, June 30, 10:30 AM – 12:00 PM

1. **The History and Current Status of Traceability** page 28
Doiron, T. (National Institute of Standards and Technology)
2. **The Uncertain State of Uncertainty in Industry** page 34
Shah, D. A. (E=mc³ Solutions)
3. **Industrial Applications of Uncertainty Analysis** page 40
Arenberg, J. W. (Northrop Grumman Space Technology)

Session III

Calibration and Measurement Modeling

Wednesday, June 30, 1:30 PM – 3:00 PM

1. **Overview of Uncertainty Analyses for Gauge Block Calibration (Invited Paper)** page 46
Decker, J. E. (National Research Council of Canada); and Schödel, R.; Bönsch, G. (Physikalisch-Technische Bundesanstalt (PTB))
2. **In-situ Calibration of a Redundant Measurement System for Manipulator Positioning** page 50
Meyer, P. J.; van Eijk, J. (Philips CFT)
3. **A Systematic Approach to the Modeling of Measurements for Uncertainty Analysis** page 56
Sommer, K.-D. (Landesamt für Mess- und Eichwesen Thüringen (LMET)); Siebert, B. R. L.; Kochsiek, M. (Physikalisch-Technische Bundesanstalt (PTB)); and Weckenmann, A. A. (University Erlangen-Nürnberg)

Session IV

Monte Carlo Methods for Uncertainty Evaluation

Wednesday, June 30, 3:30 PM – 4:30 PM

1. **Application of Simulation Software to Coordinate Measurement Uncertainty Evaluation** page 62
Summerhays, K. D.; Baldwin, J. M.; Campbell, D. A.; Henke, R. P. (MetroSage LLC)
2. **Structural Inference for Uncertainty Evaluation with Application to Calibration Experiments** page 68
Iyer, H. K.; Wang, C. M. J. (National Institute of Standards and Technology)

Session V

Uncertainty and Optical Metrology

Thursday, July 1, 8:30 AM – 10:00 AM

1. **Surface Figure Measurement at 80K: Alignment and Uncertainty Analysis** page 72
Blake, P. N.; Mink, R. G.; Chambers, J.; Davila, P. (NASA/Goddard Space Flight Center); and Robinson, F. D. (Orbital Sciences Corporation)

-
- | | | |
|----|---|---------|
| 2. | Radius of Curvature Uncertainty: Nonlinear Measurand and Treatment | page 78 |
| | Schmitz, T. L. (University of Florida); and Davies, A. D. (University of North Carolina at Charlotte) | |
| 3. | Accuracy Evaluation of White Light Interferometer Reference Signal | page 83 |
| | Wan, D.-S.; Frey, E.; Novak, E. L. (Veeco Instruments, Inc.) | |

Session VI

Interferometry/ Interlaboratory Comparisons

Thursday, July 1, 10:30 AM – 12:00 PM

- | | | |
|----|---|----------|
| 1. | Displacement Measuring Interferometry Measurement Uncertainty | page 89 |
| | Holmes, M. L.; Evans, C. J. (Zygo Corporation) | |
| 2. | Bayesian Methods for Statistical Inference on the Common Mean of Multiple Data Sources | page 95 |
| | Lu, Z.-Q. J.; Guthrie, R. W. (National Institute of Standards and Technology) | |
| 3. | Uncertainty Analysis of Interlaboratory Studies with Linear Trends | Page 100 |
| | Zhang, N. F.; Liu, H.-K.; Sedransk, N.; Strawderman, W. E. (National Institute of Standards and Technology) | |

Session VII

Applications and Measurement Modeling

Thursday, July 1, 1:30 PM – 3:30 PM

- | | | |
|----|---|----------|
| 1. | Uncertainty Analysis of Kolsky Bar Strain Gage Output * | Page 106 |
| | Rhorer, R. L. (National Institute of Standards and Technology) | |
| 2. | Characterization of a Turbine Airfoil Cooling Hole Check Standard for 5-axis CMMs –
Uncertainty Reduction by Redundancy and Error Reversal | Page 107 |
| | Drescher, J. D. (UTC – Pratt & Whitney) | |
| 3. | Using Homogenous Transformation Matrices to Define the Measurand for Uncertainty Analysis
of Complex Measurement Systems | page 113 |
| | Schmitz, T. L.; Ziegert, J. C.; Canning, J. S. (University of Florida) | |

Poster Session

Wednesday, June 20, 1:00 PM – 1:30 PM

- | | | |
|----|---|----------|
| 1. | Methods for Improving Accuracy and Repeatability in the Position of a Workpiece in Chucking | page 122 |
| | Byun, J.; Liu, C. R. (Purdue University) | |
| 2. | Cylindrical and Spherical Coordinate Transformations in Error Budgets | page 128 |
| | Smith, P. T. U. (University of Kentucky); and Vallance, R. R. (The George Washington University) | |
| 3. | Evaluation of the Geometrical Uncertainty of Helix Deviation Measurements Using
the Monte Carlo Simulation | page 134 |
| | Tanaka, H. (National Metrology Institute of Japan/AIST); Shimojo, M. (The University of
Electro-Communications); and Sato, T. (Osaka Seimitsu Kikai Co., Ltd.) | |
| 4. | The “Virtual CMM”, a Software Tool for Uncertainty Evaluation – Practical Application
in an Accredited Calibration Lab | page 140 |
| | Trenk, M. (Feinmess GmbH & Co., KG); and Franke, M.; Schwenke, H. I.
(Physikalisch-Technische Bundesanstalt (PTB)) | |
| 5. | A Framework for Uncertainty Evaluation of GPS Standard-chain | page 146 |
| | Wang, J. X.; Ma, L. M.; Jiang, X. Q.; Xu, Z. G.; Li, Z. (Huazhong University of Science & Technology) | |

* Extended abstract not available at press time

Technical Papers



A S P E

The Joint Committee for Guides in Metrology: Supplementing the Guide to the Expression of Uncertainty in Measurement

W. T. Estler
Precision Engineering Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

The ISO *Guide to the Expression of Uncertainty in Measurement*, now commonly referred to as the GUM, was developed by the ISO Technical Advisory Group 4 (TAG4) and first published in 1993.

The actual writing was done by Dr. Barry Taylor of the NIST Physics Laboratory, assisted by Dr. Chris Kuyatt from NIST's Radiation Physics Laboratory. The GUM rather quickly became a *de facto* standard procedure for the evaluation of uncertainties in measurements performed by National Metrology Institutes (NMIs). At NIST in particular, the GUM became Institute policy in the form of NIST Technical Note 1297, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*.

In addition to its use by the NMIs, the GUM has been increasingly adopted as a *de facto* standard procedure by calibration laboratories (see, for example ISO 17025), and is working its way towards application by industry down to the level of shop floor metrology. In the United States, the GUM has been formally adopted as a US National Standard in the form of ANSI/NCSL Z540-2-1997, *U.S. Guide to the Expression of Uncertainty in Measurement*.

In 1997, following the disbanding of ISO TAG4, the International Committee for Weights and Measures (CIPM) created a Joint Committee for Guides in Metrology (JCGM), chaired by Dr. Terry Quinn, then Director of the BIPM. The Committee had the task of assuming the duties of TAG4, which had developed in addition to the GUM, the *International Vocabulary of Basic and General Terms in Metrology* (known as the VIM).

The JCGM is composed, as was TAG4, of the BIPM together with the International Electrotechnical Commission (IEC), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Organization for Standardization (ISO), the International Union of Pure and Applied Chemistry (IUPAC), the International Union of Pure and Applied Physics (IUPAP), the International Organization of Legal Metrology (OIML), and the International Laboratory Accreditation Cooperation (ILAC). Additional members include representatives from NIST, PTB (Germany), NPL (United Kingdom), IMGC (Italy), and NIM (China).

Within JCGM two Working Groups were established:

- Working Group 1 (WG1), "Expression of Uncertainty in Measurement", has the task to promote the use of the GUM and to prepare supplemental guides for its broad application.
- Working Group 2 (WG2), "Working Group on International Vocabulary of Basic and General Terms in Metrology (VIM)", has the task to revise and promote the use of the VIM.

[A complete roster of the members of both Working Groups, together with descriptions of their activities, can be found at the BIPM website www.bipm.fr.]

The technical work of JCGM-WG1 is currently focused on three supplementary documents that complement and enhance the GUM in practical application. These documents are:

- A supplement to the GUM providing guidance on the use of computer-based Monte Carlo methods of propagating probability distributions for the evaluation of measurement uncertainty. Numerical methods can provide more reasonable uncertainty intervals where the measurement model is highly non-linear and/or various input quantities have asymmetric probability distributions.

A draft document, *GUM Supplement 1 - Numerical Methods for the Propagation of Distributions*, is currently being circulated for comments at the NMIs and other member organizations of WG1.

- A supplement to the GUM generalizing the propagation of uncertainty to multivariate models where there is more than one measurand. An example of such a model is a least-squares fit of experimental data to a straight line. Here the two measurands are the slope and intercept of the best-fit line, calculated from the common set of data. A draft version of this supplement is nearing completion.
- A supplement describing the role of measurement uncertainty in conformance testing and risk analysis. A draft document has been written, describing a generic approach to deciding conformance or non-conformance with specifications or requirements. Based on probability theory as the fundamental way for dealing with uncertainty, this work will yield a uniform way to treat conformance issues ranging from workpiece conformance inspections and measuring instrument verifications to the certification of standard reference materials and blood tests for the presence of chemical substances.

The GUM was published by ISO in the name of the seven contributing organizations. Since ISO owns the copyright, the GUM is available only at a price. The goal of JCGM and BIPM is to make the GUM Supplemental Guides freely available, preferably at the BIPM website. Various avenues towards this goal are being explored.

At some point in the future, a revised version of the GUM might be desirable in order to make it more fully consistent with the underlying principles of probabilistic inference. No such revision is under active consideration at the present time.

Measurement Uncertainty and Traceability Issues: A Standards Activity Update

Dr. Steven D. Phillips
Steven.Phillips@NIST.gov
Precision Engineering Division
National Institute of Standards and Technology
Gaithersburg, Maryland 20899

Introduction

In recent years, considerable interest has developed over issues concerning measurement uncertainty and traceability. The motivations for this are many. The globalization of the economy allows industry to outsource workpiece production and inspection on a worldwide scale. Hence component interchangeability (not only between components produced by one supplier but also between the same nominal component produced by several different suppliers) can only be assured only if all suppliers employ metrology to a common set of units (typically the SI units). Similarly, inspection services are frequently outsourced and the magnitude of the measurement uncertainty is taken as a measure of the quality and reliability of the measurement result; so measurement uncertainty is becoming a currency of metrology. Concomitantly, various national and international quality standards and laboratory accreditation programs are being revised to include language addressing measurement uncertainty and traceability. Finally, as workpiece tolerances steadily decrease, the cost of inspection usually increases, thus the ability to easily achieve a 10:1 ratio of the tolerance interval to measurement uncertainty interval is increasingly difficult or impossible. Measurement uncertainty is affecting the economics of production both through the cost of expensive equipment and facilities to perform metrology and in the cost of imperfect decisions, e.g., rejecting conforming workpieces or accepting nonconforming ones. Hence optimizing measurement uncertainty in an economic sense is now becoming an important issue in industry. In response to all of these factors the ASME B89.7 committee was formed to provide some standards and guidance on these topics. This paper discusses some of the current and future work of this and other standards committees.

Ten years ago the “Guide to the Expression of Uncertainty in Measurement” (GUM) [1] was published by the ISO. It is now the definitive document on evaluating measurement uncertainty. It is remarkably self-consistent and complete and has been adopted by National Measurement Institutes (NMIs), including NIST in the United States. In 1997 the GUM was adopted as a national (ANSI) standard in the US and is designated NCSL Z540-2-1997. Today the application of measurement uncertainty is going well beyond the domain of NMIs and is reaching the shop floor of industrial manufacturing enterprises. This new realm of applications is requiring both extensions of the GUM to industrial metrology situations and new documents to provide guidance and standardized practices. This paper discusses some of the recent and potential future developments on this topic.

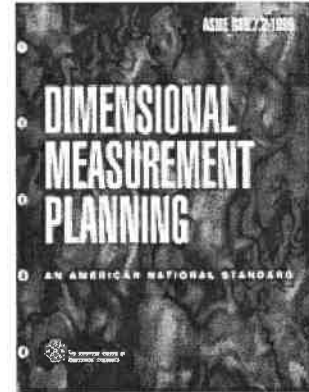


Resources for Evaluating Measurement Uncertainty

In an effort to increase the available knowledge base and provide standardized techniques for uncertainty evaluations, numerous national and international standards are under development. In the US the ASME B89.7 committee was formed to provide some guidance on these topics, with emphasis on industrial (shop floor) metrology issues. The remainder of this paper discusses some of the current and future work of this and other standards committees.

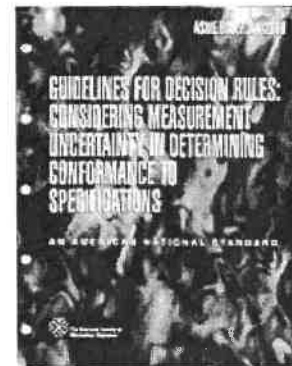
B89.7.2 (1999)

The B89.7.2 Standard “Dimensional Measurement Planning” [2] is the first published document from the B89.7 series. As the name implies, this standard is an overview of the entire measurement process. The formal part of the standard is just a brief three pages long, essentially a list of factors to consider when developing measurement plans. The bulk of the document consists of appendices that include a worked example and supporting information. B89.7.2 is a high level document. It starts by asking what measurements need to be performed, why they are being performed (e.g. process control), and what fraction (lot sampling) of the workpieces needs to be measured. Consideration is then given to selection of the measuring equipment, development of the measurement strategy, and calculation of the measurement uncertainty. Issues such as the probability of rejecting a workpiece that is within specifications or accepting one that is out of specification are also presented. Additional factors such as operator skill, the location of the measurements, measurement cycle time, and some documentary requirements are also discussed. Do not expect this standard to provide detailed information on how to make measurements or how to calculate measurement uncertainty; these are left for other more specific documents. B89.7.2 is a unique standard from the perspective of addressing the entire dimensional measurement process yielding a concise list of requirements for measurement planning.



B89.7.3.1 (2001) & ISO 14253-1 (1998)

Published in the March of 2002, the B89.7.3.1 [3] document “Guidelines for Decision Rules: Considering Measurement Uncertainty in Determining Conformance to Specifications” specifically addresses the issue of applying measurement uncertainty in industrial settings. A decision rule is a prescription for the acceptance or rejection of products based on the measurement result of a characteristic of the product, the permissible variation associated with that characteristic, and the uncertainty of the measurement result. For workpieces the permissible variation is commonly called the tolerance; for instruments it is often given by the specification limits or a maximum permissible error (MPE). We will adopt the terminology of ISO 14253-1 and refer to the permitted variation of a product’s characteristic as the specification zone.



Some measurements, particularly at NMIs, state a description of the measurement, its result, and its uncertainty; decision rules are not involved since there are no product specifications. However, most industrial measurements are performed to determine if a product is in accordance with some specification, e.g., if a workpiece is within its specified tolerance. In this situation, the measurement value is usually used in a binary decision that the product is acceptable or not

acceptable. This general class of problems, determining if a measurement result yields an acceptable product when clouded by measurement uncertainty, is not addressed by the GUM and represents an important economic (and potentially conflict prone) application of measurement uncertainty.

The concept of a decision rule has a long history and over the years it has developed many variations including “gauge maker’s rule,” “test accuracy ratio” (TAR), “test uncertainty ratio” (TUR), “four-to-one rule,” “gauging ratio,” “guard bands,” “gauging limit,” and many more. Most of these terms were defined before the development of the GUM and hence concepts such as “accuracy” or “uncertainty” were nebulously defined.

The goals of the B89.7.3.1 document are to establish a set of requirements for a decision rule, define a terminology that allows unambiguous communication of what decision rule is being used, and provide some well-documented decision rules that can be referenced.

Briefly stated, a decision rule must meet four conditions: (1) A decision rule must have a well-documented method of determining the location of the acceptance, rejection, and any transition zones (transition zones are optionally defined regions between acceptance or rejection; see Figure 4). (2) Each zone of a decision rule must correspond to a documented decision that will be implemented should the result of a measurement lie in that zone. While this is implicit for the acceptance and rejection zones by definition, any transition zones must have their corresponding decision outcomes documented. (3) A decision rule must state the procedure for addressing repeated measurements of the same characteristic on the same workpiece or instrument. (4) A decision rule must state the procedure regarding data rejection, that is, rejection of “outliers.”

The most common form of acceptance and rejection used in industry is the descendant of the “four-to-one rule” given in MIL-STD 45662A. In the new terminology this is called *simple 4:1 acceptance*. This requires that the magnitude of the expanded ($k = 2$) measurement uncertainty interval ($\pm U$) is no larger than the 1/4 of the specification zone (hence the expanded uncertainty, U , is to be no larger than 1/8 of the specification zone), and that product is accepted if the measurement result lies in the specification zone and rejected otherwise (see Figure 1).

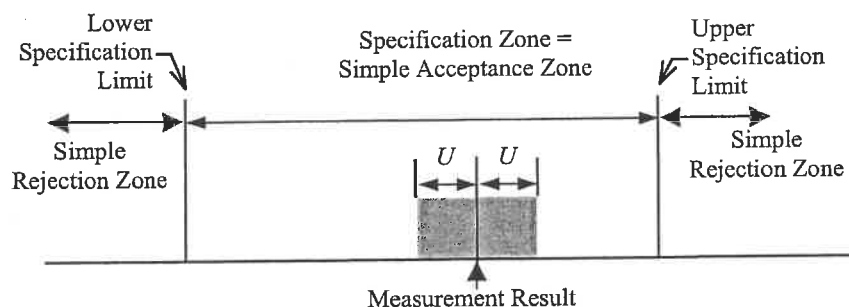


Figure 1. An example of a simple 4:1 acceptance decision rule. The measurement uncertainty interval is of width $2U$, where U is the expanded uncertainty, and the uncertainty interval is no larger than one-fourth the product’s specification zone. The measurement value shown results in product acceptance.

The simple acceptance decision rule, while straightforward, has difficulties near the specification zone limits. Due to measurement uncertainty, a product with a measurement result just inside the specification limit may actually be nonconforming. If accepting nonconforming parts has a large negative economic impact, then implementing guard banding is preferred. Guard banding is a technique that can produce a *stringent acceptance zone* that is smaller than the specification zone due to the guard bands (see Figure 2).

The size of the guard band, g , is expressed as a percentage of the expanded uncertainty, e.g. a 100 % guard band has a magnitude equal to the expanded uncertainty. Establishing the magnitude of a guard band is a business decision and is based on economics, whereas evaluating the measurement uncertainty, U , is a technical activity that depends on the measurement process.

Descriptors such as “stringent” and “relaxed,” used in describing conformance and non-conformance, have been carefully chosen. For example, stringent acceptance implies both a *decrease* in the acceptance zone width and an *increase* in confidence that a measurement result in this zone is associated with an in-specification product. Similarly stringent rejection results in a decreased size of the rejection zone while increasing the confidence that a measurement result in this zone is associated with an out-of-specification product. The converse situation applies to relaxed acceptance and rejection.

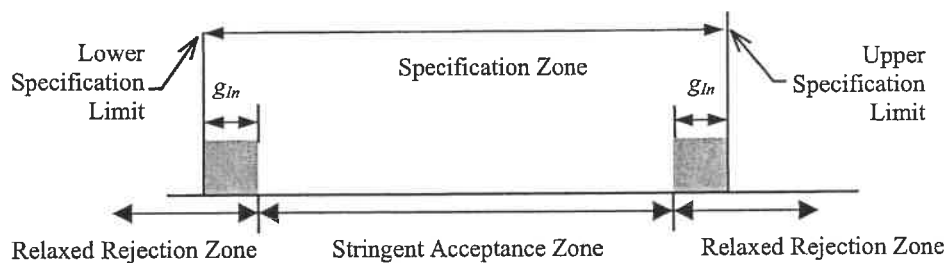


Figure 2. Stringent acceptance and relaxed rejection using symmetric two-sided guard banding. Products are accepted if the measurement result is within the acceptance zone.

If the product costs are very high, and the cost of accepting a nonconforming product is low, then *relaxed acceptance* may be preferred, see Figure 3. Relaxed acceptance allows an acceptance zone that is larger than the specification zone. This is a useful rule when a design requirement has resulted in a specification zone comparable to the state-of-the-art measurement uncertainty and hence even simple acceptance will result in a large number of conforming products being rejected.

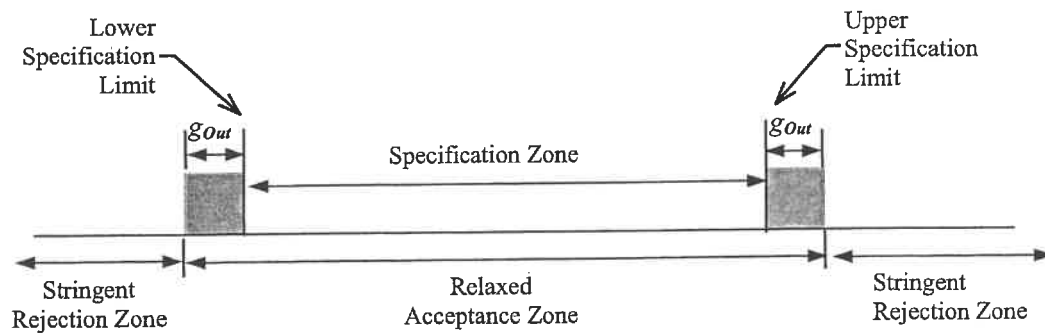


Figure 3: Relaxed acceptance and stringent rejection using symmetric two-sided guard banding. Products are accepted if the measurement result is within the acceptance zone.

Other decision rules are also possible. Figure 4 shows a situation with stringent acceptance, simple rejection, and a transition zone where the product is likely to be in conformance but the confidence of this statement is lower than that for measurement results in the stringent acceptance zone. The outcome of a measurement result in the transition zone could be, for example, selling the product at a reduced price. Ultimately the selection of a particular decision rule is a business decision that is economically driven; some of the factors to be considered are outlined in appendices of the B89.7.3.1 document.

The B89.7.3.1 document is similar to the ISO standard 14253-1 [4]. The ISO 14253-1 document focuses on the case of using stringent acceptance with a 100 % guard band for the supplier of a product and stringent rejection with a 100 % guard band for a customer seeking to reject a product. The B89.7.3 working group believes that the selection of a decision rule is a business decision, and the flexibility of having a continuum of rules ranging from stringent to relaxed acceptance or rejection is needed in order to satisfy a broad range of industries. (In B89.7.3.1, guard bands can have any percentage of the expanded uncertainty appropriate for the economics of that measurement, and it provides the terminology to communicate the type of decision rule.) Additionally, The B89.7.3.1 document establishes the requirements of a decision rule such as repeated measurements, data rejection, and documented decision outcomes, all of which are not addressed in the ISO standard.

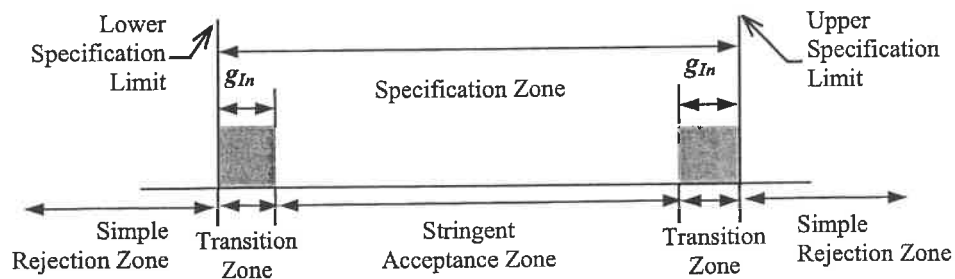


Figure 4. Stringent acceptance, simple rejection and a transition zone example using symmetric two-sided guard banding. Products are accepted if the measurement result is within the acceptance zone, rejected if in the rejection zone, and subject to a different rule in the transition zone.

ISO 14253-2 (1999)

The “Guide to the Estimation of Uncertainty in GPS Measurement in Calibration of Measuring Equipment and in Product Verification” [5] is a technical report about evaluating measurement uncertainty. The strengths of this document include a list of sources of uncertainty common to dimensional measurements and some information on evaluating these uncertainty sources.

Additionally the report describes the “Procedure for Uncertainty Management” (PUMA) method of approaching uncertainty budgets. This method suggests that the first iteration of an uncertainty budget should be a rough estimate that overestimates the uncertainty by assigning relatively large values to the uncertainty contributors and lumping many uncertainty sources together as a single contributor (i.e. input quantity). The resulting uncertainty is compared against the required application. If the evaluated uncertainty is too large, e.g. the corresponding decision rule rejects too many products, then a second iteration of the uncertainty budget is performed. The second iteration may involve both a reduction in the magnitude of various uncertainty contributors, e.g. by more careful investigation into the measurement system, and perhaps a more detailed uncertainty budget. The process is repeated until the evaluated uncertainty is sufficient for the application or the uncertainty does not decrease with additional iterations, indicating another measurement method is required.

Unfortunately, 14253-2 is also filled with quite a lot of jargon and unnecessary terminology such as “true uncertainty,” “conventional true uncertainty,” “approximated uncertainty,” and the GUM advises against using such terms. There are also philosophical differences between the GUM and the PUMA method as the GUM clearly cautions against knowingly overestimating the measurement uncertainty; nevertheless, the PUMA method is useful for industrial settings where the uncertainty of a workpiece is unlikely to be propagated into another measurement system.

B89.7.3.3 (2002) & ISO 14253-3 (2002)

B89.7.3.3 “Guidelines For Assessing the Reliability of Dimensional Measurement Uncertainty Statements” [6] is a report designed to allow parties avoid potential, or resolve actual, disagreements over the magnitude of a stated measurement uncertainty, particularly when that uncertainty is part of determining the conformance of a product to dimensional specification. With significant economic interests at stake, it is not surprising that customers and suppliers might disagree over the magnitude of the measurement uncertainty statement. Applying these guidelines can assist businesses in avoiding disagreements about measurement uncertainty statements between customers and suppliers and in resolving such disagreements should they occur. Disagreements over uncertainty statements involving a single measurement system and multiple measurement systems (each having its own uncertainty statement) are considered. Guidance is provided for examining uncertainty budgets as the primary method of assessing their reliability. Additionally, resolution by direct measurement of the measurand is also discussed. While the document was initially designed for resolving disagreements over measurement uncertainty, the report discusses many factors of formulating uncertainty budgets that will be useful to anyone responsible for this task.

The ISO 14253-3 [7] document, like the B89.7.3, is concerned with resolving disagreements between two parties over a dispute of a measurement uncertainty statement. The ISO document

tends toward a more formal flowchart approach than its US counterpart, which is more tutorial in focus. The main goal of ISO 14253-3 is achieving agreement between the parties whereas B89.7.3 emphasizes the metrological issues seeking to achieve agreement through education.

Future B89.7 documents

The B89.7 committee has several ongoing work items. The B89.7.4 working group has completed a draft of a report that addresses the quantitative risk assessment of decision rules. This document provides mathematical guidance for determining the fraction of conforming parts rejected and nonconforming parts accepted for various decision rules under various assumptions of the production and measurement uncertainty probability distributions. This report is anticipated to be available in 2004.

Also under development is B89.7.3.2, which examines a simplified GUM approach to measurement uncertainty. Finally, the B89.7.8 working group is considering issues associated with measurement traceability and how to achieve it. This document will discuss the differences between traceability and measurement uncertainty and the steps needed to provide the required documentation.

Other Recent Documents Available On Line

Fortunately, several documents on measurement uncertainty are available for free and can be downloaded from the Internet. A brief description of each document is presented and a URL address given.

NIST "Technical Note 1297" [8] is sometimes called a summary of the GUM; it contains the basic definitions and method. While the GUM is highly recommended, TN 1297 is available at no cost and hence could be freely distributed within an organization.

URL: <http://physics.nist.gov/Document/tn1297.pdf>

An online introduction to the GUM and the SI system of units is available on NIST's web site.

URL: <http://physics.nist.gov/cuu/Uncertainty/index.html>

"A Careful Consideration of the Calibration Concept" [9] is primarily focused on uncertainty issues related to calibrations, i.e. measurements that result in the issuing of certificates describing the accuracy of the measurement. The paper does contain an introduction useful for all metrologists in defining the measurement under consideration, i.e. the measurand. This often-overlooked factor can result in protracted arguments between suppliers and customers, for example a supplier might measure the diameter of a bore as acceptable using a plug gauge while the customer rejects this feature using a least-squares fit on a coordinate measuring machine. No amount of improvement in the accuracy of these measuring methods will resolve this discrepancy as two fundamentally different measurands are under inspection. The document also has an extensive appendix that is a tutorial on basic issues of uncertainty such as the distinction between measurement uncertainty and error.

URL: <http://nvl.nist.gov/pub/nistpubs/jres/106/2/j62phi.pdf>

"Uncertainty and Dimensional Calibrations" [10] provides an excellent discussion of sources of uncertainty relevant to the calibration of dimensional artifacts and gauges. A generic uncertainty

budget is presented and nine examples, including gauge blocks, ring gauges, optical flats, and sieves are worked out in detail. While the intended audience is gauge calibration laboratories, all metrologists will benefit from the clear presentation and application of the GUM to several different measurement situations.

URL: <http://nvl.nist.gov/pub/nistpubs/jres/102/6/j26doi.pdf>

For the advanced measurement uncertainty expert the following papers may be of interest.

“The Calculation of Measurement Uncertainty using Prior Information” [11] discusses a Bayesian inference approach to including prior information about the value of the measurand in the calculation of measurement uncertainty;

URL: <http://nvl.nist.gov/pub/nistpubs/jres/103/6/j36phi.pdf>

“Guidelines for Expressing the Uncertainty of Measurement Results Containing Uncorrected Bias” extends the GUM to the case of including uncorrected systematic errors in an expanded measurement uncertainty statement.

URL: <http://nvl.nist.gov/pub/nistpubs/jres/102/5/j25phi.pdf>

“A Distribution-Independent Bound of the Level of Confidence in the Result of a Measurement” [12] discusses the relationship between the coverage factor in the expanded uncertainty and the corresponding level of confidence.

URL: <http://nvl.nist.gov/pub/nistpubs/jres/102/5/j25est.pdf>

Finally, a good source of additional publications on measurement uncertainty can be found at the Bureau International des Poids et Mesures (BIPM) website of the Joint Committees for Guides in Metrology.

URL: http://www.bipm.org/CC/documents/JCGM/bibliography_on_uncertainty.html

Acknowledgements

The author would like to thank Dr. Tyler Estler, Mr. Bruce Borchardt, and Dr. Craig Shakarji, for their review of the manuscript. This work was supported in part by the Air Force’s Coordinated Calibration Group, and the Manufacturing Engineering Laboratory of NIST.

References

- 1 International Organization for Standardization, "Guide to the Expression of Uncertainty in Measurement," Geneva, Switzerland, 1993. (corrected and reprinted 1995). This document is also available as a U.S. National Standard: NCSL Z540-2-1997.
- 2 ASME B89.7.2-1999 “Dimensional Measurement Planning,” available from the American Society of Mechanical Engineers. www.asme.org.
- 3 ASME B89.7.3.1-2001 “Guidelines for Decision Rules: Considering Measurement Uncertainty in Determining Conformance to Specifications,” available from the American Society of Mechanical Engineers. www.asme.org.

-
- 4 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 1: Decision rules for providing conformance or non-conformance with specification," ISO 14253-1 (1998).
 - 5 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 2: Guide to the Estimation of Uncertainty in GPS Measurement in Calibration of Measuring Equipment and in Product Verification," ISO 14253-2 (1999).
 - 6 ASME B89.7.3.3-2002 "Guidelines For Assessing the Reliability of Dimensional Measurement Uncertainty Statements," available from the American Society of Mechanical Engineers; www.asme.org.
 - 7 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 3: Guidelines to achieving agreements on measurement uncertainty statements," ISO 14253-3 (2002).
 - 8 "Guidelines for Evaluation and Expressing the Uncertainty of NIST Measurement Results," B.N. Taylor and C.E. Kuyatt, NIST Technical Note 1297, 1994 Edition, National Institute of Standards and Technology, Gaithersburg, MD 20899, 1994.
 - 9 "A Careful Consideration of the Calibration Concept," S.D. Phillips, W.T. Estler, T. Doiron, K.R. Eberhardt, M.S. Levenson", *Journal of Research of NIST*, **106**, 1 – 9 (2001).
 - 10 "Uncertainty and Dimensional Calibrations," T. Doiron and J. Stoup, *Journal of Research of NIST*, **102**, 647-676, (1997).
 - 11 "Calculation of Measurement Uncertainty Using Prior Information," S.D. Phillips, and W.T. Estler, *Journal of Research of NIST*, **103**, 625-632 (1998).
 - 12 "A Distribution-Independent Bound of the Level of Confidence in the Result of a Measurement," W.T. Estler, *Journal of Research of NIST*, **102**, 587-588 (1997).

Uncertainty due to Finite Resolution Measurements

S.D. Phillips, B. Tolman, T.W. Estler
National Institute of Standards and Technology
Gaithersburg, MD 20899
Steven.Phillips@NIST.gov

Abstract

We investigate the influence of finite resolution on measurement uncertainty from a perspective of the *Guide to the Expression of Uncertainty in Measurement* (GUM). Finite resolution in the presence of Gaussian noise yields a distribution of results that strongly depends on the relative location of the true value to the resolution increment. We show that although there is no simple expression relating the standard deviation of the distribution to the uncertainty at a specified level of confidence, there is an analytic relation between the mean value and the standard deviation. We further investigate the conflict between the GUM and ISO 14253-2 regarding the method of evaluating the standard uncertainty due to finite resolution and show that, on average, the GUM method is superior, but still approximate.

1 Introduction

The issue of measurement uncertainty is becoming increasingly relevant to both calibration laboratories and factory floor metrology. In some cases measurement uncertainty has significant economic impact on the cost of a product. For example, the ISO standard 14253-1 [1] (default condition) requires the expanded uncertainty to be subtracted from both ends of the product tolerance yielding a smaller “conformance zone.” A measurement result must lie in this zone in order for the manufacturer to distribute the product. Avoiding over estimation of the measurement uncertainty can result in a larger conformance zone and hence lower product cost. In some situations the finite resolution of a measuring instrument is a significant contributor to the uncertainty statement. For example, the manufacturers of hand held digital calipers typically set the resolution of the instrument such that repeated measurements of the same quantity yields nearly the same result, within one or two units of the least count (i.e. resolution) of the instrument. (In principle the caliper would be more accurate with an additional display digit, however, customers perceive quality as the ability of the instrument to yield the same value for repeated measurements.) The uncertainty budget of such hand held instruments typically contains only a few significant influence quantities since most quantities such as the accuracy of the calibration standards (e.g. gauge blocks) and thermal effects are usually small compared to the instrument resolution. Therefore the inclusion (or not) of the uncertainty associated with the finite resolution of the display can significantly affect the magnitude of the uncertainty statement. Similar issues may arise with the finite resolution introduced by analog to digital conversion electronics where repeated sampling of a signal differs by only one or two bits. In this paper we first examine some general properties of the probability distribution associated with finite resolution and then examine how it impacts the uncertainty evaluation under different measurement scenarios.

The Guide to the Expression of Uncertainty in Measurement (GUM) [2] identifies the finite resolution of a measuring instrument as a contributor to measurement uncertainty and suggests that it should be evaluated by a (Type B) uniform distribution with a full width of one resolution

unit resulting in a standard uncertainty of $\frac{1}{\sqrt{12}}$ in units of the resolution (GUM F.2.2.1). The GUM also suggests that this contribution is uncorrelated to other uncertainty sources and that it should be added in a root-sum-of-squares (RSS) manner in the uncertainty statement (e.g., see GUM H.6).

Alternatively, ISO 14253-2 [3] recommends examining the standard uncertainty of the resolution relative to the standard deviation of repeated measurements and then including the larger of the two in the uncertainty evaluation and discarding the lesser of the two. The rationale of this method is that the resolution is already intertwined with the standard deviation of repeated measurements since that data is recorded with the resolution of the instrument.

We summarize these two procedures as:

Rule 1: RSS the standard uncertainty of the resolution (Type B via a uniform distribution) with the uncertainty associated with the standard deviation of repeated measurements in the uncertainty evaluation.

Rule 2: Include the larger of the following two: the standard uncertainty of the resolution (Type B via a uniform distribution) and the standard deviation of repeated measurements in the uncertainty evaluation, and discard the lesser of the two.

Clearly Rule 1 and Rule 2 cannot both simultaneously be the best estimate of the uncertainty for all measurement cases.

There are two measurement scenarios we will consider. The “special test” scenario involves constructing an uncertainty statement for one specific measurement quantity. Typically this will involve repeated observations of the quantity, each recorded with finite resolution. The best estimate of the measurand is considered as the mean of the repeated observations (after all corrections are applied) and the uncertainty statement will be associated with this mean value. The “measurement process” scenario involves constructing an uncertainty statement that will be applicable to a series of future measurement results. This is typical of commercial metrology where a large number of nearly identical artifacts or workpieces are produced and only a single measurement, having finite resolution, is recorded. In this scenario, the uncertainty statement is developed once and then associated with each future (single observation) measurement result. The uncertainty evaluation process will involve, among other things, repeated observations recorded with finite resolution and used to characterize the measurement variation.

Without loss of generality we consider the case where a resolution increment has the value of unity, and the infinite resolution value of the measurement, μ , lies between zero and one-half (all other cases are modulo this problem). We consider the case where the measurement is corrupted by noise from a Gaussian distribution with standard deviation σ . The situation of interest will be for $\sigma < 1$ since large σ is equivalent to infinite resolution. Our approach will be to examine both the special test scenario and the measurement process scenario in the limit of large sample size.

2 Probability Distributions

In the limit of large sample size, statistics can be reasonably estimated by their corresponding expectation values. Finite resolution requires that only integer values can be observed. Hence a continuous probability distribution function (pdf) becomes a discrete pdf that can be written:

$$\text{Discrete pdf}(x) = w(\mu, \sigma, n) \delta(x - n) \quad n: -\infty, \dots, -3, -2, -1, 0, 1, 2, 3, \dots, +\infty$$

$$\text{with } w(\mu, \sigma, n) = \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \text{pdf}(\mu, \sigma, x) dx = \Phi(\mu, \sigma, n + \frac{1}{2}) - \Phi(\mu, \sigma, n - \frac{1}{2}) \quad (1)$$

where w is a weight function, $\delta(x - n)$ is a delta function, and Φ is the cumulative probability distribution function. Figure 1 illustrates the quantization (using equation 1) of a Gaussian distribution with $\mu = 0.4$ and $\sigma = 1$.

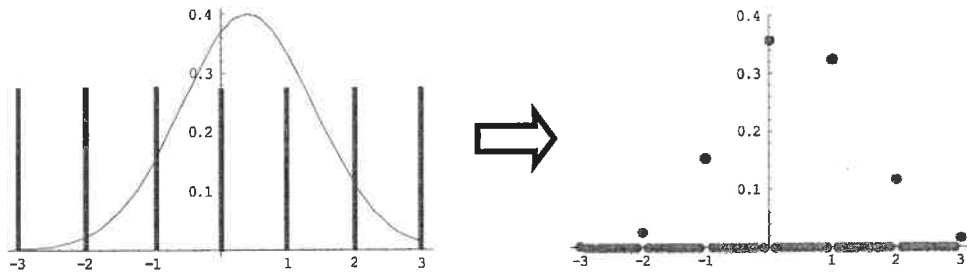


Figure 1. A Gaussian distribution with $\mu = 0.4$ and $\sigma = 1$ is converted into a discrete distribution.

3 Descriptive Statistics: Mean and Standard Deviation

Using the discrete probability distribution, expectation values can be calculated. The expected value of the sample mean, $\bar{x}$, and the sample standard deviation, s , can be calculated from:

$$\bar{x}(\mu, \sigma) = \sum_{n=-\infty}^{+\infty} w(\mu, \sigma, n) n \quad \text{and} \quad s(\mu, \sigma) = \sqrt{\sum_{n=-\infty}^{+\infty} w(\mu, \sigma, n) (n - \bar{x})^2} \quad (2)$$

Figure 2 displays plots of $\bar{x}$ and s for relevant values of μ and σ . The plot for $\bar{x}$ shows the expected step function (due to the discrete resolution) at $\mu = 0.5$ when $\sigma \rightarrow 0$, and the expected behavior $\bar{x} \rightarrow \mu$ for large σ . Similarly, the plot for s displays the expected rapid rise when $\mu \rightarrow 0.5$ and σ is small and the expected behavior $s \rightarrow \sigma$ for large σ .

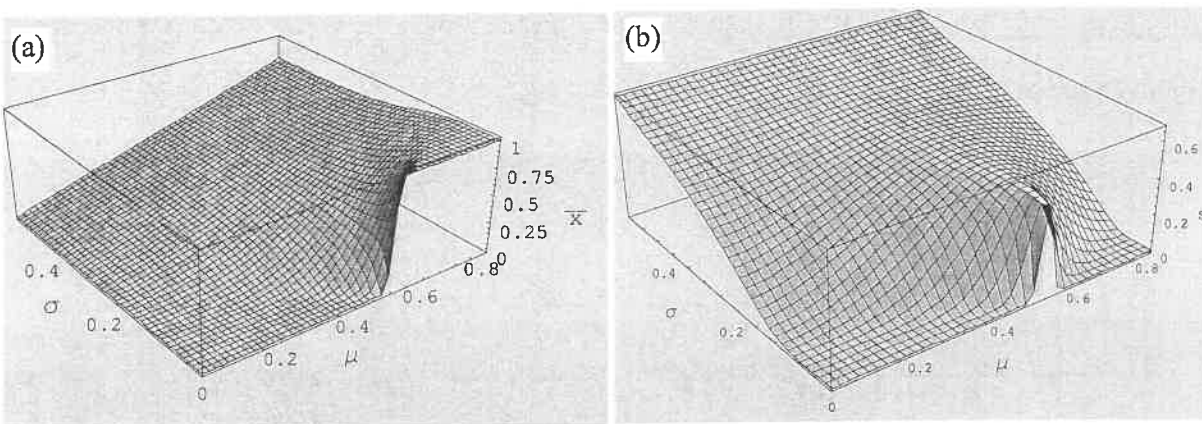


Figure 2. Plots of (a) the sample mean, $\bar{x}$, and (b) sample standard deviation, s , as a function of μ and σ for data recorded with unit resolution.

An interesting feature of Figure 2 can be more easily seen in Figure 3 which shows the sample mean $\bar{x}$ as a function of σ for various values of μ . For any given value of $\bar{x} \leq 0.5$, which is a horizontal line on Figure 3, the possible values of μ are always greater than $\bar{x}$. Hence the observed sample mean is always biased toward the closest resolution increment regardless of the value of σ (for large σ this bias becomes insignificant). For small σ , e.g. $\sigma < 0.5$, it is incorrect to believe that uncertainty associated with the sample mean approaches zero as a result of averaging a large sample; rather it approaches a fixed value resulting in a systematic error.

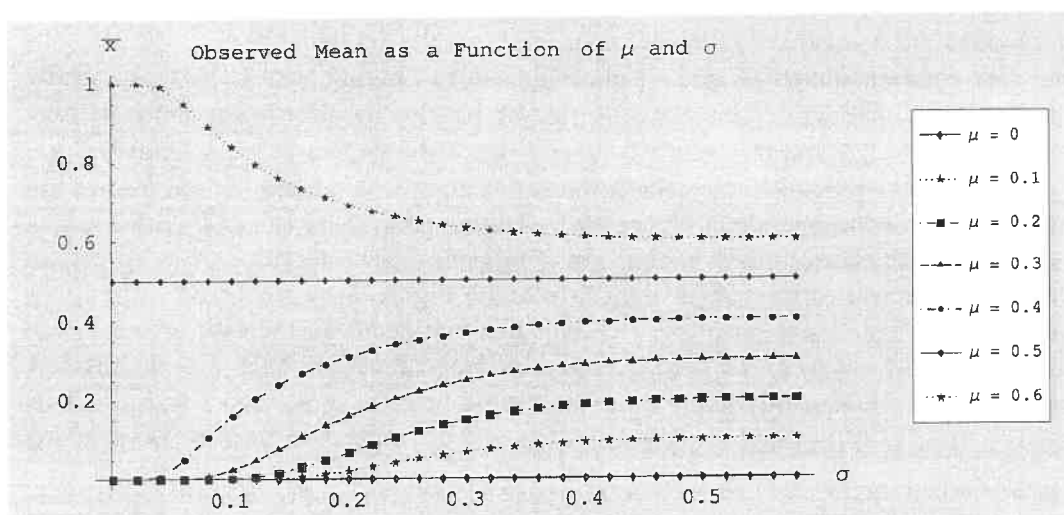


Figure 3. Plot of $\bar{x}$ vs. σ for various values of μ , note that for any particular $\bar{x}$ the possible values of μ are always greater than $\bar{x}$ resulting in a bias of $\bar{x}$ toward the resolution unit.

Figure 4 shows the dependence of s on σ for various values of μ . As σ becomes large, $s \rightarrow \sigma$, approaching it from above. The slight over estimation of σ by s is well known and the difference

is called Sheppard's correction [4]. We further point out that $s \geq \sigma$ when $s > \frac{1}{\sqrt{12}}$ regardless of μ . Similarly, $\frac{1}{\sqrt{12}} \geq \sigma$ when $s < \frac{1}{\sqrt{12}}$ regardless of μ ; hence $\text{Max}[\frac{1}{\sqrt{12}}, s] \geq \sigma$ for all μ ; a fact we will make use of later.

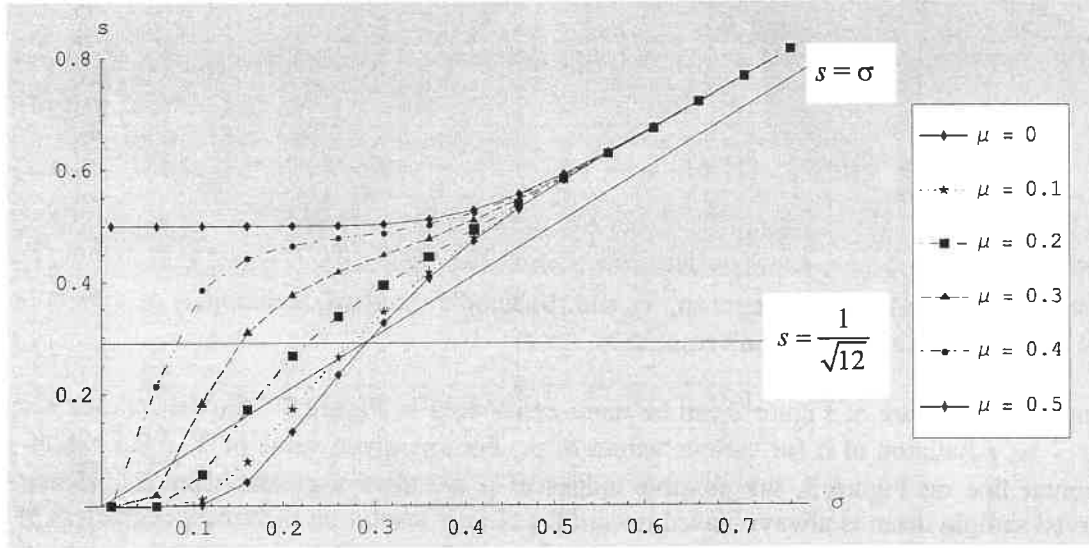


Figure 4: Plot of the standard deviation of the finite resolution population as a function of the standard deviation of the Gaussian noise.

What is not obvious from examining the previous plots is the interaction of $\bar{x}$ and s . In particular, there are combinations of $\bar{x}$ and s that are forbidden. Figure 5(a) shows a plot of the $\bar{x}$, s space generated (numerically, with a sample size of 10,000) by allowing μ and σ to vary over the domain $0 \leq \mu \leq 0.5$ and $0 \leq \sigma \leq 0.5$, where the Z dimension is the systematic (i.e. expected) error $\bar{x} - \mu$. As previously described, the expected error is always biased toward the resolution unit, i.e. the bias is negative in Figure 5(a). Additionally, there is a disk with a radius of one-half resolution unit, as described in equation (3), inside which no value of $\bar{x}$ and s can occur. This disk is more clearly seen in the 2D plot of Figure 5(b) where only the $\bar{x}$, s coordinates are shown. The dense set of $\bar{x}$, s coordinates that occur on the disk boundary are understood by considering values of n where the equality in equation (3) holds. For the discrete pdf given by equation (1) the equality holds when the weight functions $w(0, \mu, \sigma) + w(1, \mu, \sigma) = 1$. That is, for any μ ($0 \leq \mu \leq 0.5$) and σ such that $\Phi(\mu, \sigma, 1\frac{1}{2}) - \Phi(\mu, \sigma, -\frac{1}{2}) = 1$ then all of the probability is contained in the weight functions $w(0, \mu, \sigma)$ and $w(1, \mu, \sigma)$ and this forces the equality of equation (3) and consequently the value of $\bar{x}$ and s lie on a circular arc of radius one half resolution unit. (Since all Gaussian distributions have some probability, albeit infinitesimal, to extend to infinity, the disk radius is actually a limiting value.) As seen in Figure 5(a), it is only the values where $\left(\frac{1}{2} - \bar{x}\right)^2 + s^2 = \left(\frac{1}{2}\right)^2$ that are significantly biased while other values of $\bar{x}$, s away from this disk edge rapidly become an unbiased estimator of μ .

$$\left(\frac{1}{2} - \bar{x}\right)^2 + s^2 = \left(\frac{1}{2}\right)^2 + \sum_{n=-\infty}^{n=+\infty} w(n, \mu, \sigma)(n^2 - n) \geq \left(\frac{1}{2}\right)^2 \quad \forall n, \mu, \sigma \quad (3)$$

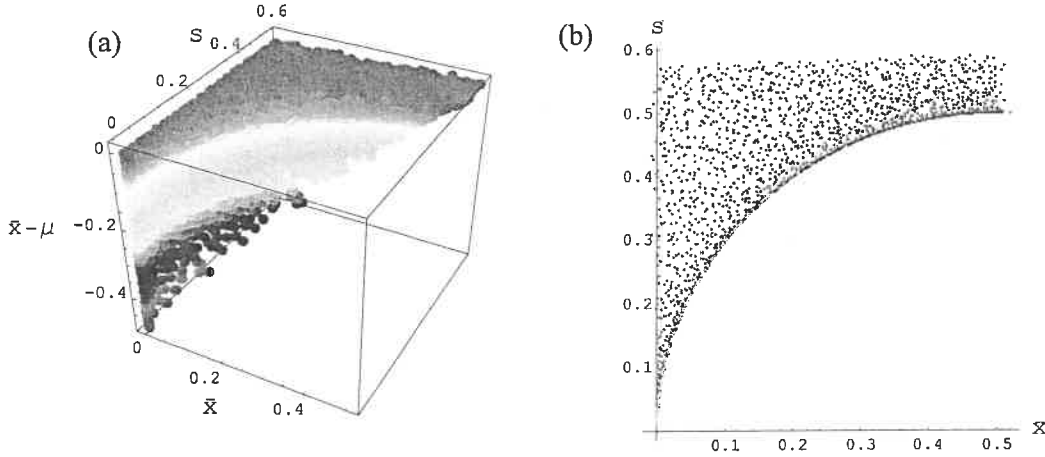


Figure 5: (a) the value of $\bar{x}$, s , and $\bar{x} - \mu$ for all combinations of $0 \leq \mu \leq 0.5$ and $0 \leq \sigma \leq 0.5$; (b) the 2D plot of the $\bar{x}$, s coordinates shown in (a).

4 Errors and Level of Confidence

For the discrete pdf given by equation (1) with standard deviation given by equation (2) we can explore the relationship between the standard deviation and the level of confidence, i.e. the coverage factor. Although we are considering random Gaussian noise corrupting our measurement result, the subsequent rounding due to finite resolution significantly changes the level of confidence for a coverage factor of $k = 2$, particularly when $\sigma < 0.5$. We can gain some insight into this effect by examining the magnitude of the error at the 95th percentile in relation to the magnitude of the standard deviation. The magnitude of the error that occurs at the 95th percentile is given by the cumulative probability distribution associated with equation (1) and evaluated to determine the integer, $n_{95\%}$, and the corresponding error, $\varepsilon_{95\%}$, by increasing the number of terms in equation (4) until the inequality just holds.

$$w(0, \mu, \sigma) + w(1, \mu, \sigma) + w(-1, \mu, \sigma) + w(2, \mu, \sigma) + \dots + w(n_{95\%}, \mu, \sigma) \geq 0.95 \quad (4)$$

$$\varepsilon_{95\%} = |n_{95\%} - \mu|$$

Clearly $\varepsilon_{95\%}$ will depend on the particular values of μ and σ and we expect that $\varepsilon_{95\%}$ will also display the discreteness of the underlying discrete pdf. For example, as σ increases with μ fixed, $\varepsilon_{95\%}$ will remain fixed until the inequality in equation (4) is violated forcing the addition of another term to the cumulative probability and a corresponding jump in the value of $\varepsilon_{95\%}$. Figure 6(a) shows the $\varepsilon_{95\%}$ value over the region: $0 \leq \mu \leq 0.5$ and $0 \leq \sigma \leq 1$. Figure 6(b) displays the required coverage factor to achieve a 95% level of confidence; i.e. $k = \varepsilon_{95\%} / s$. For μ small but nonzero and σ small, the coverage factor approaches infinity since all the measured values of x yield zero resulting in $s \rightarrow 0$ while $x - \mu$ remains finite. One method to avoid infinite coverage

factors is to establish a finite lower bound for the standard uncertainty; this is employed by Rules 1 and 2.

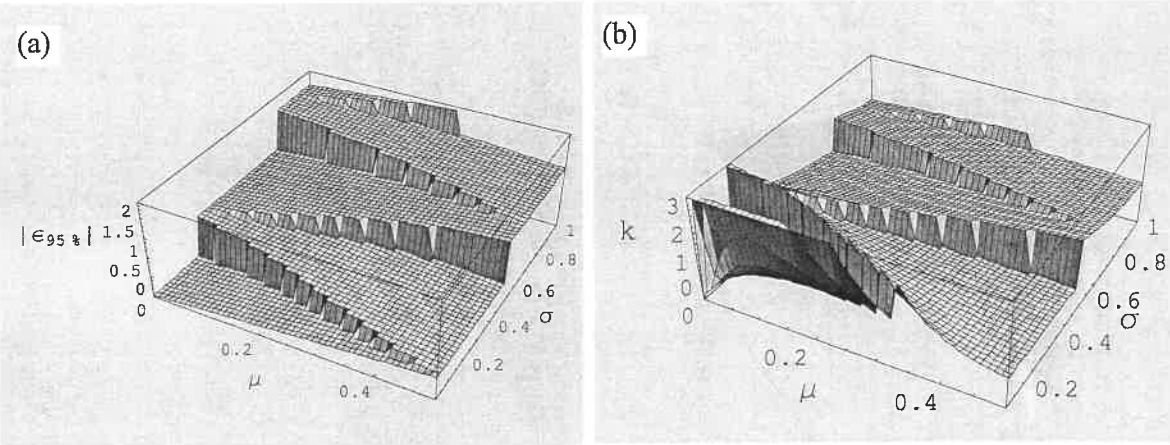


Figure 6: (a) The 95th percentile error (in units of resolution) as a function of the true value (μ) and the standard deviation of the Gaussian noise (σ); (b) the corresponding coverage factor required to include the 95th percentile error

In the special test scenario many measurements are taken on the object under consideration and the average value (which in general will not be an integer) is computed as the best estimate of the measurand. The uncertainty of the mean value is of interest, however the mean value has a systematic error, i.e. the expectation value of $(\bar{x} - \mu)$, that is biased toward the resolution unit for values of $\bar{x}$, s that lie on the forbidden disk edge (see Figure 5), the details of this bias is shown in Figure 7(a). Also shown in Figure 7(a) is the lower bound of the expected error given by $\bar{x} - \frac{1}{2}$, note that the range of errors for a given $\bar{x}$ result from different μ , σ points mapping to the same value of $\bar{x}$, but having different systematic errors. Figure 7(b) shows the points in μ , σ space that map to values of $\bar{x}$, s that lie on the disk boundary; this represents a significant region of μ , σ space and thus gives rise to the density of $\bar{x}$, s points on the disk boundary in shown in Figure 5(b). The bold points shown in Figure 7(b) all map very close to the point $\bar{x} = 0$, $s = 0$.

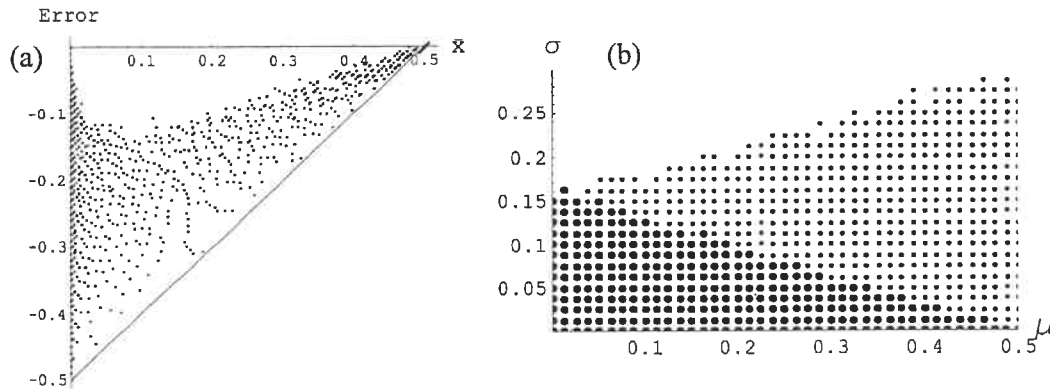


Figure 7: (a) the error ($\bar{x} - \mu$) for points on the disk boundary shown in Figure 5. (b) the points in μ , σ space that map to the disk boundary; the bold points map to $\bar{x} = 0$, $s = 0$.

It is tempting to apply a correction for the average systematic error (as a function of $\bar{x}$) to obtain a better estimate of μ . However, since the bias is nearly a step function that rapidly approaches zero away from the disk edge, applying such a correction based on the proximity of a $\bar{x}$, s coordinate to the disk edge is subject to misapplication due to statistical variations expected in any finite sample size. A simpler but cruder approximation would be to use the uncorrected value of $\bar{x}$ and consider the systematic error to be bounded between zero and the line $\bar{x} - 0.5$ whenever $s \leq 0.6$. We will consider an uncertainty rule based on this approximation for the special test scenario.

5 Measurement Process Scenario

In the measurement process scenario an uncertainty evaluation is performed on a measurement system and then a future (single observation) measurement result, x , which represents the best estimate of the measurand is assigned an uncertainty based on the prior evaluation. Since x is a single reading from an instrument having resolution R , and in this paper we take $R = 1$, hence x must be an integer. One complicating factor in this scenario is that the reproducibility evaluation is conducted on a reference artifact having some value μ_{Ref} while the uncertainty statement is relevant to a future measurement having a different (infinite resolution) value μ ; and the two values are independent. Hence all of the measurements used in the uncertainty evaluation to determine s , (based on a large sample of measurements) will depend on the particular value of μ_{Ref} , and different values of μ_{Ref} will yield different values of s ; see Figure 2b. (Note that the best estimate of the reference artifact, $\bar{x}_{\text{Ref}}$, is discarded since it provides no information about the value of the future measurement result). In this paper we will examine only two values for μ_{Ref} . The case $\mu_{\text{Ref}} = 0$ is common in practice since reference artifacts used in uncertainty evaluations usually have a value that is exactly on a unit of resolution. For example, a digital caliper reading in units of inches that is evaluated by examining repeated measurements on gauge blocks from an English (inch based) set. The caliper may have a resolution of 0.001 inch and the gauge blocks will be within a few microinches of some multiple of 0.001 inch. (Note for this condition to hold the systematic error of the instrument must be small with respect to the resolution.) Alternatively, we examine the case where μ_{Ref} is uniformly distributed between 0 and 0.5, denoted $\Sigma\mu$. This could occur when repeated measurements are made on several different metric gauge blocks using an inch based caliper and the results are pooled to obtain the standard deviation.

We consider three rules to evaluate the standard uncertainty due to finite resolution and repeatability in the measurement process scenario given in equation (5);

$$\text{Rule 1: } u(x) = \sqrt{\left(\frac{\frac{1}{2}R}{\sqrt{3}}\right)^2 + s(\mu_{\text{Ref}})^2}$$

$$\text{Rule 2: } u(x) = \text{Max} \left[\frac{1}{\sqrt{3}} R, s(\mu_{\text{Ref}}) \right] \quad (5)$$

$$\text{Rule 3: } u(x) = \sqrt{\left(\frac{1}{\sqrt{3}} R \right)^2 + \left(\text{Max} \left[\frac{1}{\sqrt{3}} R, s(\mu_{\text{Ref}}) \right] \right)^2}$$

where R is the resolution (in this paper $R = 1$), and $s(\mu_{\text{Ref}})$ is the standard deviation of a large sample of measurements on an artifact with a (infinite resolution) value μ_{Ref} .

Rules 1 and 2 are based on the GUM and ISO 14253-2 respectively as previously described. Rule 3 includes the observation that $\text{Max}[\frac{1}{\sqrt{12}}, s] \geq \sigma$. All three rules contain two quantities, one dependent on R and the other on s , and a prescription to combine them. The first quantity (with $R = 1$) has the value $\frac{1}{\sqrt{12}}$ and represents the standard uncertainty associated with a uniform

distribution with limits $\pm \frac{1}{2}$ representing that the single (future) observation x could be located anywhere in the resolution interval. The second quantity is an estimate of the underlying population variance σ that gives rise to the observed variation in x . That is, the first quantity accounts for the ambiguity due to the resolution and the second quantity accounts for the reproducibility of the measurement result. Rule 1 always uses the calculated standard deviation as an estimate for σ . However, when σ is small, s can underestimate σ , e.g. if $\mu = 0$ and $\sigma = 0.2$ nearly all of the repeated observations will be zero and hence $s \approx 0$. Rule 2 also estimates σ using s and then selects the larger of s or the standard uncertainty of the resolution. We can expect that by omitting one of the uncertainty sources that Rule 2 will underestimate the uncertainty associated with the measurand. We introduce Rule 3 that estimates σ by selecting a quantity that is an upper bound for σ , as previously noted in section 3.

6 Testing the Uncertainty Rules for the Measurement Process Scenario

We can estimate how effective these rules are in containing the reasonable values that can be attributed to the measurand. In the measurement process scenario, the single measurement result, x , is determined by the value of μ (which is a property of the object but is unknown), corrupted by a random perturbation arising from σ , and rounded to unit resolution. We can imagine that this measurement is repeated a large number of times and then ask what fraction of the errors are contained in the uncertainty statement, where the error is $\varepsilon = x - \mu$, and is due to the random Gaussian perturbation and the rounding to an integer value. We will adopt this viewpoint since in commercial metrology, it is typical to construct an expanded uncertainty using the coverage factor $k = 2$ and then expect that 95 % of the potential measurement errors to be contained within this interval.

One complicating factor in the measurement process scenario is that the probability of a future measurement result is contained within the expanded ($k = 2$) uncertainty interval is a function of

three population variables: μ_{Ref} , μ , and σ . Our approach will be to fix the population variables, evaluate $U_{k=2}$, and then draw a large number of samples (each representing a future measurement) and determine what fraction are contained within the expanded uncertainty interval. We will then select another set of population variables and continue this methodology until we have examined the set of population variables of interest. Figure 8 displays the probability that a future measurement result is contained within the expanded uncertainty interval evaluated of Rule 1 (the GUM rule) with $\mu_{\text{Ref}} = 0$ over the domain $0 \leq \mu \leq 0.5$ and $0 < \sigma \leq 1$. As seen in the figure, the surface consists of a relatively flat mesa having a containment probability near unity, interrupted by abrupt canyons of significantly lower containment probability. The corresponding plots using Rule 2 and Rule 3 are similar, but with broader - deeper canyons, and narrower - shallower canyons, respectively. We note that Figure 8, and the associated three statistics in Table 1, that the containment probability only addresses the question of what fraction of errors (for a particular μ , σ) are contained in the uncertainty interval, it does not address their relative position within the interval. Hence the same containment probability would be assigned to two different distributions having the same fraction of uncontained errors, despite that one distribution might have the uncontained errors only slightly away from the uncertainty interval and the other may have its uncontained errors grossly away from the uncertainty interval.

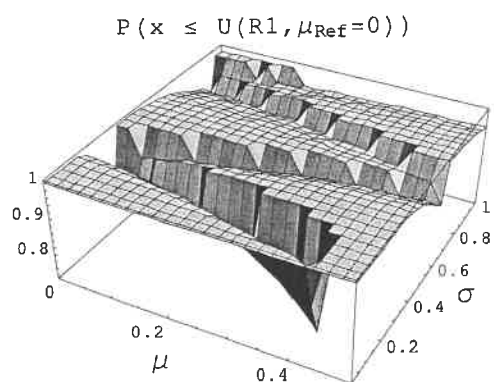


Figure 8: A plot of the probability that the measurement error of a single observation (having value μ and corrupted by Gaussian noise of standard deviation σ) will be contained within the expanded uncertainty of Rule 1 evaluated with $\mu_{\text{Ref}} = 0$.

Table 1 attempts to summarize these results. Min P is the minimum average containment probability over the domain of population variables examined. That is, over the domain of μ , σ values considered, find the particular value of μ , σ that has the smallest containment probability and report that value as Min P. For example, when Rule 1 is evaluated with $\mu_{\text{Ref}} = 0$ the minimum average containment probability occurs at $\sigma = 0.15$ and $\mu = 0.4$; that is, for this set of values a (randomly chosen) future measurement result has only a 0.75 probability of being contained in the expanded uncertainty interval of this rule. The Min P value tells us something about the lowest point on the surface shown in Figure 8.

Avg P is the average probability of a measurement error containment over all values of μ , σ considered; for the previous example of Rule 1 with $\mu_{\text{Ref}} = 0$, Avg $P(x \leq U) = 0.97$, hence the probability of error containment for most values of μ and σ is near unity.

Most metrologists expect approximately 95% of the potential measurement errors to be contained in the expanded ($k = 2$) uncertainty statement. However, even for the infinite resolution case of a Gaussian population 95% is the expected fraction of error containment, hence any finite sample size will have slightly more or less than 95% of the deviations from the mean contained within $2s$. As another measure of the effectiveness of an uncertainty rule we introduce the percentage of the surface of Figure 8 that does not include at least a 94% containment probability, this is denoted $\% P \leq 0.94$; the value of 0.94 was selected since, for these sample sizes, the corresponding infinite resolution case always included at least this fraction of the errors. Note both the $\% P \leq 0.94$ and $\text{Avg } P(x \leq U)$ depend on the size μ , σ domain under consideration, hence they are only relative metrics that are useful when comparing rules over the same μ , σ domain.

Table 1 Measurement Process Scenario

	Rule 1 (GUM)		Rule 2 (ISO)		Rule 3		∞ Resolution $U = 2s$
	$\mu_{\text{Ref}} = 0$	$\mu_{\text{Ref}} = \Sigma\mu$	$\mu_{\text{Ref}} = 0$	$\mu_{\text{Ref}} = \Sigma\mu$	$\mu_{\text{Ref}} = 0$	$\mu_{\text{Ref}} = \Sigma\mu$	
Min $P(x \leq U)$	0.75	0.84	0.65	0.74	0.85	0.91	0.94
Avg $P(x \leq U)$	0.97	0.98	0.95	0.97	0.98	0.98	0.95
$\% P \leq 0.94$	13 %	7%	30 %	19 %	10 %	5%	0.00
Avg $(U - \epsilon_{95})$	0.23	0.29	0.12	0.16	0.27	0.30	0.02
Max $(U - \epsilon_{95})$	0.69	0.89	0.58	0.76	0.81	0.89	0.07
Min $(U - \epsilon_{95})$	-0.13	-0.25	-0.38	-0.48	-0.15	-0.14	0.00

From Table 1 we see that none of the three (ad hoc) rules guarantee 95% error containment for all values of μ and σ , with Rule 2 failing this criteria 30 % of the time for the $\mu_{\text{Ref}} = 0$ case. All of the three rules benefited from computing the standard deviation of the measurement process using many reference standards with different values and pooling the results, i.e. the $\mu_{\text{Ref}} = \Sigma\mu$ case. Indeed, for the $\mu_{\text{Ref}} = \Sigma\mu$ case, Rule 1 (GUM) performance approached that of Rule 3 without the need to compare the relative size of s to the standard uncertainty of the resolution, as required in both Rule 2 and Rule 3.

Also shown in Table 1 are three statistics that describe the difference between the expanded uncertainty and the magnitude of the 95th percentile error, $|\epsilon_{95\%}|$. This tells us something about the amount of the over (or under) estimation of the uncertainty interval. As might be expected for simple ad-hoc uncertainty rules, none of the three rules adapt themselves well to the discontinuous nature of the discrete pdf, e.g. see Figure 6, and at some values of μ , σ significantly overestimate the magnitude of $|\epsilon_{95\%}|$ while at other points they underestimate its magnitude. These statistics will be of more value for the special test measurement scenario.

7 Special Test Scenario

In the special test scenario, a large sample of repeated observations is used to calculate both $\bar{x}$ and s . Hence, in this scenario the number of observations used to calculate $\bar{x}$ is the same as the number of observations used to calculate the standard deviation s . In the special test scenario the GUM rule (Rule 1) combines the uncertainty of the resolution unit with the standard deviation of the sample mean. For large samples, the uncertainty associated with the resolution unit will clearly dominant since the standard deviation of the mean scales as $1/\sqrt{N}$. Similarly, the ISO

rule (Rule 2) selects the larger of the two terms; hence we expect these two rules to be similar since the uncertainty of the resolution unit dominates the standard deviation of the mean in both cases. Consequently, it is clear that for large sample sizes both these rules overestimate the uncertainty associated with $\bar{x}$.

For Rule 3 we use our observation that near the forbidden disk region (rather crudely described by $s \leq 0.6$) that $\bar{x}$ is a biased estimator of μ and the value of μ is contained in the interval $\bar{x} \leq \mu \leq 0.5$ as seen in Figure 3. While in principle a correction should be applied to account for this bias, in practice the typical user of a hand held instrument is unlikely to accommodate such an inconvenience. Consequently we describe the ignorance in the location of μ by a uniform distribution as seen in Rule 3 of equation (7). Furthermore, Rule 3 includes the observation that $\text{Max}[\frac{1}{\sqrt{12}}, s] \geq \sigma$; outside the disk region Rule 3 treats the uncertainty evaluation as the infinite resolution case.

$$\begin{aligned}
 \text{Rule 1: } u(\bar{x}) &= \sqrt{\left(\frac{\frac{1}{2}R}{\sqrt{3}}\right)^2 + \frac{(s)^2}{N_{\bar{x}}}} \\
 \text{Rule 2: } u(\bar{x}) &= \text{Max} \left[\frac{\frac{1}{2}R}{\sqrt{3}}, \frac{s}{\sqrt{N_{\bar{x}}}} \right] \\
 \text{Rule 3: } u(\bar{x}) &= \sqrt{\left(\frac{\frac{1}{2}R - \bar{x}}{\sqrt{3}}\right)^2 + \frac{\left(\text{Max} \left[\frac{\frac{1}{2}R}{\sqrt{3}}, s \right]\right)^2}{N_{\bar{x}}}} \quad \text{if } s \leq 0.6R \\
 &= \frac{s}{\sqrt{N_{\bar{x}}}} \quad \text{if } s > 0.6R
 \end{aligned} \tag{7}$$

8 Testing the Uncertainty Rules for the Special Test Scenario

In the case of a large sample size both the value of $\bar{x}$ and s become relatively stable and consequently the magnitude of the systematic error $|\bar{x} - \mu|$ and the uncertainty rule become stable. Hence we can consider $U - |\bar{x} - \mu|$ over a region of μ, σ values to examine if a particular rule contains (or not) the systematic error.

Table 2 displays the information regarding the containment probability for the three rules as applied to the special test scenario. Both Rule 1 and Rule 2 greatly overestimate the uncertainty associated with the mean and hence always contain 100 % of the errors. Rule 3 more closely

approaches the 95% error containment interpretation with a minimum containment probability of 0.94 and an average (over the μ, σ values considered) of 0.98.

To consider the magnitude of the overestimation we examine the expected values of the difference between the expanded uncertainty and $|\epsilon_{95\%}|$. As previously suggested, Rule 1 and Rule 2 behave similarly and both typically overestimate $|\epsilon_{95\%}|$ by one-half a resolution unit. Rule 3 does significantly better by more closely matching the expanded uncertainty to the $|\epsilon_{95\%}|$ value, with an average overestimation of 0.15 resolution units.

Table 2: Special Test Scenario

	Rule 1 (GUM)	Rule 2 (ISO)	Rule 3
Min $P(x \leq U)$	1.0	1.0	0.94
Avg $P(x \leq U)$	1.0	1.0	0.98
% $P \leq 0.94$	0	0	1%
Avg $U - \epsilon_{95} $	0.51	0.50	0.15
Max $U - \epsilon_{95} $	0.58	0.58	0.58
Min $U - \epsilon_{95} $	0.18	0.18	-0.01

In the special test scenario with a large sample size, Rule 1 and Rule 2 become identical with

$$u(\bar{x}) \rightarrow \frac{1}{\sqrt{12}}, \text{ and Rule 3 yields } u(\bar{x}) \rightarrow \frac{\frac{1}{2} - \bar{x}}{\sqrt{3}} \text{ if } s < 0.6 \text{ and } u(\bar{x}) \rightarrow 0 \text{ if } s \geq 0.6. \text{ Elster [5]}$$

has examined the general solution for the special test scenario using Bayesian calculations. He also considered the special case of data with an equal number of ones and zeros (hence $\bar{x} = 0.5$ and $s = 0.25$), and showed for large samples that the best estimate of μ was the mean $\bar{x} = \frac{1}{2}$ and $u(\bar{x}) \rightarrow 0$; this result is given by Rule 3 but not by Rule 1 or Rule 2 which yield $u(\bar{x}) \rightarrow \frac{1}{\sqrt{12}}$.

9 Conclusions

We have examined the statistics associated with data recorded with finite resolution in the limit of large sample size. Due to finite resolution, there exists a disk of defined by $\left(\frac{1}{2} - \bar{x}\right)^2 + s^2 = \left(\frac{1}{2}\right)^2$ inside which the values of $\bar{x}, s$ are forbidden. Furthermore, values of $\bar{x}$ that occur on the boundary of the disk are always (except for the $\mu = 0$ and $\mu = 0.5$ case) systematically biased toward the resolution unit. Away from the disk boundary $\bar{x}$ rapidly becomes an unbiased estimator of μ .

We have also considered various simple ad-hoc rules for evaluating the uncertainty due to finite resolution including those recommended for the by the GUM and ISO 14253-2. For the measurement process scenario and using the criteria that the $(k = 2)$ expanded uncertainty should include approximately 95 % of the potential errors, we found that the ISO 14253-2 rule underestimates the uncertainty for roughly 30 % of the μ, σ values, while the GUM rule and Rule 3 do significantly better. Additionally, we pointed out that evaluating the standard deviation of repeated measurements using several different reference artifacts, with values that

are not increments of the resolution unit, improves the containment probability all of the uncertainty rules examined.

For the special test scenario, where a large number of measurements are available to compute the mean value and the standard deviation, both the GUM and ISO 14253-2 rules greatly overestimate the uncertainty of the mean for large sample sizes. This overestimation results from setting a lower limit for the uncertainty equal to $\frac{1}{\sqrt{12}}$ regardless of sample size or the value of the standard deviation. Rule 3 more closely matches the expanded uncertainty to the 95th percentile error while still maintaining a 95% error containment probability.

10 References

- 1 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 1: Decision rules for providing conformance or non-conformance with specification," ISO 14253-1 (1998).
- 2 International Organization for Standardization, "Guide to the Expression of Uncertainty in Measurement," Geneva, Switzerland, 1993. (corrected and reprinted 1995). This document is also available as a U.S. National Standard: NCSL Z540-2-1997.
- 3 International Standard, "Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring instruments -- Part 2: Guide to the Estimation of Uncertainty in GPS Measurement in Calibration of Measuring Equipment and in Product Verification," ISO 14253-2 (1999).
- 4 Stuart, A. and Ord, J.K., Kendall's Advanced Theory of Statistics: Distribution Theory, 6th edition, New York: Oxford University Press, 1998.
- 5 Elster C., Evaluation of Measurement Uncertainty in the Presence of Combined Random and Analogue to Digital Conversion Errors, Measurement Science Technology 11, 1359-1363, (2000).

The History and Current Status of Traceability

Ted Doiron
Engineering Metrology Group
National Institute of Standards and Technology

6.10 traceability

property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.

The definition of uncertainty has changed little since it was required in Mil Std 45662A, the only change being the addition of the final 4 words. For national level laboratories, as well as scientific work in general, the additional phrase mandating uncertainty statements was of no particular note since uncertainty statements were always considered necessary. In the industrial measurement community, however, it was a bombshell that unmoored traceability from its historical practice and set it adrift. In this paper I will try to give some historical perspective to the current state of affairs, and discuss the increasing confusion about the relation of traceability to uncertainty. Finally, I will try to show that there are no real technical issues to settle because the laboratory accreditation community has already, and properly, addressed traceability as a contractual and regulatory issue.

In a scientific sense, traceability has never been an issue. Scientists never give traceability statements. They do give estimates of uncertainty, and one source of uncertainty is that the unit might not be exactly the stated unit. In other words, if a scientist states an answer is meters, everybody assumes he is using the international meter, and any uncertainty in his meter being the SI meter is somehow handled in the uncertainty budget. For the scientist, a measurement that is not traceable is inconceivable.

For industrial metrology, traceability is a critical property of measurements. The reasons are basically historical. The gradual implementation of the idea that industrial measurements should be made like scientific measurements has taken a long and somewhat twisted path. A century ago industry got along quite well with a system that looked at a factory or company as its own little universe. As long as measurements were consistent in the little universe, it was of no major concern that the different universes used slightly different units. As industry became more linked, that is, sold parts and assemblies to each other this separate universe paradigm no longer worked. The era when different industries needed linkages between local units is quite varied. Both world wars I and II caused a flurry of standardization efforts and international agreements. Still, it was not until 1959 that a meeting of the directors of the English speaking NMIs adopted a single definition of the inch, 25.4 mm. Semiconductor factories were still separate universes with respect to details like linewidth within the memory of most readers.

Traceability of Standards

The term “traceability” itself became a serious matter with the US Government adoption of quality standards in the middle of the last century. It was an age of specification writing, with detailed

specifications for everything from quality systems to doughnuts. Among these was MIL Q-9858A which stated:

4.2 Measuring and Testing Equipment.

These devices shall be calibrated against certified measurement standards which have known valid relationships to national standards at established periods to assure continued accuracy.

One thing of note is that this standard is somewhat better in its language. If the language “known relationship” had been used instead of “traceability” the system might have been closer to our current thinking on the subject.

MIL 45662A defined traceability as:

- 3.5 *Traceability. The ability to relate individual measurement results through an unbroken chain of calibrations to one or more of the following:*
 - 3.5.1 *U.S. national standards maintained by the U.S. National Bureau of Standards (NBS) and the U.S. Naval Observatory;*
 - 3.5.2 *Fundamental or natural physical constants with values assigned or accepted by the U.S. NBS;*
 - 3.5.3 *National standards of other countries which are correlated with U.S. national standards;*
 - 3.5.4 *Ratio type of calibrations;*
 - 3.5.5 *Comparison to consensus standards.*

The standard then defined the requirements for traceability in section 5.8.

5.8 Calibration sources. M&TE and measurement standards shall be calibrated by the contractor or another calibration facility utilizing measurement standards whose calibration is traceable. All measurement standards used in the calibration system shall be supported by certificates, reports, or data sheets attesting to the description or identification of the item; the calibration source; date of calibration; calibration assigned value; statement of uncertainty and environmental or other conditions under which the calibration results were obtained.

This “traceability” is not quite our current traceability in that it specifies only the traceability of equipment. Thus, the values expressed by the measurement standards used in calibrations, not the measurement results, needed to be traceable. It is from this point that the definition of traceability begins to shift.

The problem with traceability was the chain. In many cases the gage blocks on the floor were related to national standards by chains of 5 to 10 steps. There was simply no mechanism to actually go through the chain from the bottom to the top for all gages and measuring equipment. Into this vacuum stepped the NBS Number. MIL Handbook 52B was the guidance document for MIL C-45662A. It stated:

6. Reports for the highest level standards of sources other than NBS or a Government laboratory must bear a statement that comparison has been made and is traceable to

National Standards at planned intervals. An NBS test number is one means of substantiating comparison.

The “one means” eventually became THE MEANS when the individual standards, like the gage block standard GGG-G-15, called for specific items to be in a calibration report.

3.9 Manufacturer's report of calibration. A report of calibration is required (see 6.2), it shall be furnished with each gage block or set of gage blocks showing the results obtained at the manufacturer's final inspection of each individual block and containing the following information:

(g) Statement on traceability of calibration to NBS shall include NBS test number and date.

Thus the “Traceability = NBS Number” became the US norm.

Before we go on, it is important not to miss the forest for the trees, or in this case one tree. The military quality standards are an important landmark in metrology history. The entire system was designed to help assure that equipment, and in our case measurements, met a specification. MIL STD 45662A was an important step towards modern industrial metrology methods and is a direct ancestor of the current ISO 17025. Of the large requirements structure of the standard, traceability is a small part, and a rather limited part at that.

While the system might look fragile, particularly the NBS Number somehow enforcing accuracy, we are still faced with the undeniable fact that the system seemed to work. Airplanes fly, cars roll, and this laptop I am using works quite well. It was not an efficient system; some calibration labs were audited by their customers over 100 times a year. But the combination of two things seems to have been adequate for the purpose.

1. The existence of a valid NBS number on the lab calibration report was reasonable assurance that an NBS calibration was somewhere in the chain.
2. The standards used were better than what was actually needed by some significant ratio.

Traceability of Measurements

Traceability, of itself, does not make measurements better. The entire structure of the MIL standards did force better measurements, or at least placed some emphasis on better measurements. Somehow, traceability became an important issue separate from the regulatory environment that started it. By the end of the 1970s there was a cottage industry of traceability paper writing, mostly at NBS. There were problems fitting statistical process control and measurement assurance practices into the industrial environment where calibration intervals and stickers were a way of life, a way of life required by MIL C-45662A. There began to be a linkage between traceability and uncertainty. Traceability was something that had a “timeline”. One system could be “more traceable” than another. These are attributes of uncertainty, not traceability. Traceability is the existence of a chain, and thus a logical variable. There is a chain or there isn't. Still uncertainty was somehow implicitly tangled up in the picture.

specifications for everything from quality systems to doughnuts. Among these was MIL Q-9858A which stated:

4.2 Measuring and Testing Equipment.

These devices shall be calibrated against certified measurement standards which have known valid relationships to national standards at established periods to assure continued accuracy.

One thing of note is that this standard is somewhat better in its language. If the language “known relationship” had been used instead of “traceability” the system might have been closer to our current thinking on the subject.

MIL 45662A defined traceability as:

- 3.5 *Traceability. The ability to relate individual measurement results through an unbroken chain of calibrations to one or more of the following:*
 - 3.5.1 *U.S. national standards maintained by the U.S. National Bureau of Standards (NBS) and the U.S. Naval Observatory;*
 - 3.5.2 *Fundamental or natural physical constants with values assigned or accepted by the U.S. NBS;*
 - 3.5.3 *National standards of other countries which are correlated with U.S. national standards;*
 - 3.5.4 *Ratio type of calibrations;*
 - 3.5.5 *Comparison to consensus standards.*

The standard then defined the requirements for traceability in section 5.8.

5.8 Calibration sources. M&TE and measurement standards shall be calibrated by the contractor or another calibration facility utilizing measurement standards whose calibration is traceable. All measurement standards used in the calibration system shall be supported by certificates, reports, or data sheets attesting to the description or identification of the item; the calibration source; date of calibration; calibration assigned value; statement of uncertainty and environmental or other conditions under which the calibration results were obtained.

This “traceability” is not quite our current traceability in that it specifies only the traceability of equipment. Thus, the values expressed by the measurement standards used in calibrations, not the measurement results, needed to be traceable. It is from this point that the definition of traceability begins to shift.

The problem with traceability was the chain. In many cases the gage blocks on the floor were related to national standards by chains of 5 to 10 steps. There was simply no mechanism to actually go through the chain from the bottom to the top for all gages and measuring equipment. Into this vacuum stepped the NBS Number. MIL Handbook 52B was the guidance document for MIL C-45662A. It stated:

- 6. *Reports for the highest level standards of sources other than NBS or a Government laboratory must bear a statement that comparison has been made and is traceable to*

National Standards at planned intervals. An NBS test number is one means of substantiating comparison.

The “one means” eventually became THE MEANS when the individual standards, like the gage block standard GGG-G-15, called for specific items to be in a calibration report.

3.9 Manufacturer's report of calibration. A report of calibration is required (see 6.2), it shall be furnished with each gage block or set of gage blocks showing the results obtained at the manufacturer's final inspection of each individual block and containing the following information:

(g) Statement on traceability of calibration to NBS shall include NBS test number and date.

Thus the “Traceability = NBS Number” became the US norm.

Before we go on, it is important not to miss the forest for the trees, or in this case one tree. The military quality standards are an important landmark in metrology history. The entire system was designed to help assure that equipment, and in our case measurements, met a specification. MIL STD 45662A was an important step towards modern industrial metrology methods and is a direct ancestor of the current ISO 17025. Of the large requirements structure of the standard, traceability is a small part, and a rather limited part at that.

While the system might look fragile, particularly the NBS Number somehow enforcing accuracy, we are still faced with the undeniable fact that the system seemed to work. Airplanes fly, cars roll, and this laptop I am using works quite well. It was not an efficient system; some calibration labs were audited by their customers over 100 times a year. But the combination of two things seems to have been adequate for the purpose.

1. The existence of a valid NBS number on the lab calibration report was reasonable assurance that an NBS calibration was somewhere in the chain.
2. The standards used were better than what was actually needed by some significant ratio.

Traceability of Measurements

Traceability, of itself, does not make measurements better. The entire structure of the MIL standards did force better measurements, or at least placed some emphasis on better measurements. Somehow, traceability became an important issue separate from the regulatory environment that started it. By the end of the 1970s there was a cottage industry of traceability paper writing, mostly at NBS. There were problems fitting statistical process control and measurement assurance practices into the industrial environment where calibration intervals and stickers were a way of life, a way of life required by MIL C-45662A. There began to be a linkage between traceability and uncertainty. Traceability was something that had a “timeline”. One system could be “more traceable” than another. These are attributes of uncertainty, not traceability. Traceability is the existence of a chain, and thus a logical variable. There is a chain or there isn't. Still uncertainty was somehow implicitly tangled up in the picture.

The final linkage was completed in the definition in the VIM. Here, two things changed. First, traceability moved from being a property of the measurement standards to the measurement result. The second change was to demand a reported uncertainty on each calibration.

The changes in the definition of traceability, while appropriate, had the consequence of destabilizing traceability. It was now obvious that an NBS number was not adequate proof of traceability. Somehow the lab at the bottom would have to prove that all the links existed and had uncertainty statements. For a while, at least in the US, the requirement was simply ignored. The new definition of traceability was only in the International Standards system, and had not been adopted into the US Government or US voluntary standards system. This changed dramatically with the rise of laboratory accreditation in the 1990s.

Laboratory Accreditation

The US calibration system abruptly joined the rest of the world in 1994 with the adoption of NCSL/ANSI Z540-1994 and the cancellation of 45662A. This standard adopted the language of ISO Guide 25 for the US calibration quality standard. It was not a complete break with the past, Guide 25 was basically a modified version of 45662A after it had passed through NATO into the European standards system. Thus, most of the quality system requirements were not much different than the one currently in use in the US. The new definition of traceability, however, was now the only definition.

During the same period that the quality standard was changing laboratory accreditation hit the US calibration system. Laboratory accreditation, like the registration to ISO 9000 movement, can be thought of as the final formalization of supply chain audits. Now, instead of each supplier being responsible for auditing their suppliers, a third party would audit them. The direct economic benefit was that a calibration lab that had been audited by 100 or more different customers each year would now be audited once, and all of the companies would accept it. This was a significant savings for the calibration system as a whole, and since most of these audits were required by and paid for by the US government it was a savings for everybody.

While the need for laboratory accreditation can be debated, one thing it did simplify was traceability. The basic premise of accreditation is that each lab need only worry about its link in the traceability chain and the accreditation organization will assure the rest of the chain. In its simplest implementation it would mean that each lab would have to use suppliers that were accredited by the same organization. In practice, the organizations have organized under the International Laboratory Accreditation Cooperation (ILAC) to allow labs to use suppliers accredited by other organizations that have reciprocity agreements. Thus the different assessment organizations assess each other on a regular basis and the web of reciprocal organizations is large. In the US there are a number of organizations that have wide reciprocity with both the European (EA) and Pacific (APLAC) mutual recognition bodies.

In order to have reciprocity, there must be essential agreement of what the standard (ISO 17025) means, and this includes traceability. Traceability has gone from being a US government contractual term to an internationally recognized regulatory term, defined in detail by ILAC.

ILAC G2:1994 Traceability of Measurements

4. ELEMENTS OF TRACEABILITY

Traceability is characterized by a number of essential elements:

- an unbroken chain of comparisons going back to a standard acceptable to the parties, usually a national or international standard;*
- measurement uncertainty; the measurement uncertainty for each step in the traceability chain must be calculated according to defined methods and must be stated so that an overall uncertainty for the whole chain may be calculated;*
- documentation; each step in the chain must be performed according to documented and generally acknowledged procedures; the results must equally be documented;*
- competence; the laboratories or bodies performing one or more steps in the chain must supply evidence for their technical competence (e.g. by demonstrating that they are accredited);*
- reference to SI units; the “appropriate” standards must be primary standards for the realization of the SI units;*
- recalibrations; calibrations must be repeated at appropriate intervals; the length of these intervals depends on a number of variables, (e.g. uncertainty required, frequency of use, way of use, stability of the equipment).*

Widespread laboratory accreditation moves the term “traceability” completely out of the scientific arena to the international regulatory arena. The US, though a bit behind because of a late start, is catching up fast. There are over 800 accredited laboratories in the US.

It would be nice if this was the end of the story, but there are always some details to clean up. In traceability the detail is how many traceability paths must be documented when the measurement entails the measurement of a number of different quantities. To measure speed you must measure distance and time. This is not so hard. But to measure length, you must measure temperature. How well do you need to measure temperature?

If I am measuring a 1 meter piece of wood with a tape, and need an uncertainty of 1 mm, do I need a traceable temperature measurement? The worst case (across the grain) CTE wood is about $60 \times 10^{-6}/^{\circ}\text{C}$, and my tolerance means I need to know the temperature to about 20°C . I would maintain I can reliably guess the room temperature to much better than that, although I have never been personally calibrated.

At what point does having a calibrated thermometer become silly? There is a complete continuum of cases between the NMI measurements where millidegrees are important to the case above. Nearly any attempt to address all of the possible cases is a bottomless pit. The simplest thing is to say that any correction to a measurement result that is over some % of the total uncertainty must

come from a traceable measurement. The problem even with this is the % needed will vary by field, and in a regulatory environment, well meaning advice can have serious consequences. Currently, the decisions are left to the technical assessor. The lab can, of course, appeal any decision and have other experts weigh in on the subject.

The problem is well understood at the international level. The ILAC G16 document of 2001 states its policy on traceability and has the following requirement and note:

(a) Laboratories accredited by ILAC Member Bodies shall be able to demonstrate that calibration of *critical equipment*, and hence the calibration or test results generated by that equipment, relevant to their scopes of accreditation, is traceable to the International System of Units (SI units).

Note 3:

“Critical “ equipment used by testing and calibration laboratories is considered by ILAC to be those items of equipment necessary to perform a test or calibration from the scope of accreditation and which have a significant effect on the uncertainty of measurement of test or calibration results. ILAC member bodies have agreed to investigate this issue further and to develop guidelines to differentiate between calibrations that are critical and less critical and to indicate how in the latter case the traceability requirements may be less rigorous.

Thus the problem is being addressed. Finally, it should be recognized that refining the meaning of traceability is not a game that just anybody can play. Laboratory accreditation is a world-wide movement with very rigorous requirements for the member bodies to use standardized interpretations of terminology, traceability included. The terms must be usable in such diverse fields as chemical and environmental testing, ASTM materials tests, and gauge blocks. Given the breadth of uses and the international nature of its definition and interpretation, it few remaining details that are unsettled will remain unsettled for some time to come.

There is still a useful niche for professional and standards groups, which is to give advice on how to make up uncertainty budgets in their respective fields. Some fields, like dimensional metrology, have a reasonable number of references for metrologists to consult, although there is room for more. Traceability, however, is best left to the regulatory community that defines and enforces it.

References:

1. Department of Defense MIL-STD-45662A, Calibration Systems Requirements, 1 August 1988.
2. MIL-Q-9858A, “Quality Program Requirements,” Dec. 16, 1963.
3. MIL Handbook 52B, “Evaluation of Contractor’s Calibration System”, Feb. 1, 1989.
4. International Vocabulary of Basic and General Terms in Metrology, second edition, 1993, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, International Organization for Standardization, Geneva (1993).
5. National Conference of Standards Laboratories (NCSL), ANSI/NCSL Z540-1-1994, American National Standard for Calibration—Calibration Laboratories and Measuring and Test Equipment—General requirements, First Edition, August 1994, p. 7.
4. ILAC-G2:1994 “Traceability of Measurements”, 1994
5. ILAC-G16:2001, “ILAC Policy on Traceability of Measurement Results”, 2001.

The Uncertain State of Uncertainty in Industry

Dilip A. Shah

President, E = mc³ Solutions

Excellence = Measurement, Calibration, Computers, Consulting Solutions

197 Great Oaks Trail #130

Wadsworth, Ohio 44281-8215

Tel: 330-328-4400 Fax: 330-336-3974

E-mail: emc3solu@aol.com

1.0 INTRODUCTION

The ISO 17025 laboratory accreditation standard requires measurement uncertainty estimates of calibration and test processes for the laboratory's scope of accreditation. Measurement uncertainty budget estimates are examined by accrediting body assessors before accreditation is granted to a laboratory. Because of the requirement by the international standard, the laboratories are compelled/forced/dragged to understand the process of estimating measurement uncertainty as outlined in the Guide to Uncertainty of Measurement (herein referred to as GUM).

It is important to note the reliance of industry and the world commerce on good measurements. Without good measurements, the everyday commercial activities that we take for granted would come to its knees and chaos would develop. Yet, industry in general takes good measurement practices for granted and is ignorant of measurement uncertainty associated with making a measurement. Improving technology and innovation continues the drive towards better measurement instrumentation at less cost than the previous versions. Adding to the confusion are old vestiges of 10:1 or more recently, 4:1 Test Accuracy Ratios (herein referred to as TAR) requirements spelled out in calibration contracts, old standards and still practiced/required by the industry.

Too often the end-user of the equipment focuses on getting the measurement equipment calibrated for another year and having a sticker on the equipment attesting to that fact. This narrow focus misses out on an important aspect of the measurement process, the uncertainty associated with making a measurement and applying it correctly.

This paper will identify the GUM requirements for estimation of Measurement Uncertainty and how the requirements are misstated or wrongly interpreted by the industry. The paper also discusses the myths surrounding the Test Accuracy Ratios in today's calibration environment and why TARs should be relegated to its place in history.

2.0 GENERAL DEFINITIONS

The Vocabulary of International terms in Metrology (Herein referred to as VIM) defines the following:

- **Accuracy** - The closeness of agreement between a test result and the accepted reference value.
- **Bias** - The difference between the expectation of the test results and an accepted reference value.
- **Drift** is a slow change of a metrological characteristic of a measuring instrument.
- **Error** (of indication) of a measuring instrument is the indication of a measuring instrument minus a 'true' value of the corresponding input quantity, i.e. it has a sign
- **Laboratory bias** - the difference between the expectation of the test results from a particular laboratory and an accepted reference value.
- **Measurand** - particular quantity subject to measurement.
- **Measurement** - set of operations having the object of determining a value of a quantity.
- **Metrology** - the science of measurement.
- **National (measurement) standard** - standard recognized by a national decision to serve, in a country, as the basis for assigning values to other standards of the quantity concerned.
- **Precision** The closeness of agreement between independent test results obtained under stipulated conditions.

- **Resolution** - smallest change of measured quantity which changes the indication of an measuring instrument
- **Repeatability** - precision under repeatability conditions
- **Repeatability conditions** - where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.
- **Reproducibility** - precision under reproducibility conditions
- **Reproducibility conditions** - where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment
- **Stability** refers to the ability of a measuring instrument to maintain constant its metrological characteristics with time
- **Testing** is a technical investigation, e.g. as to whether a product fulfils its specified performance.
- **Traceability** means that a measured result can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties.
- **Trueness** - the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. The measure of trueness is usually expressed in terms of bias.
- **Uncertainty** of measurement is a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand. It can also be expressed as an estimate characterizing the range of values within which the true value of a measurand lies.
- **Verification** is an investigation that shows that specified requirements are fulfilled.

3.0 TEST ACCURACY RATIO (TAR) MYTH

This is the most commonly abused requirement stated in calibration and test requirements by the industry. In simple terms, 10:1 TAR states that the standard used to calibrate the equipment has to be capable of achieving 10 times the accuracy.

It is easy to acquire a micrometer capable of measuring 0.00005 inches for less than US \$ 100.00. In order to calibrate this micrometer, one would require a standard capable of measuring 0.000005 inches (implying 10:1 ratio). While this may be possible in this case, consider the other requirement for calibration, which is traceability. This means that the standard used has to have an unbroken chain of comparisons to a national standard (NIST in USA). This calibration traceability pyramid is illustrated in Figure 1.

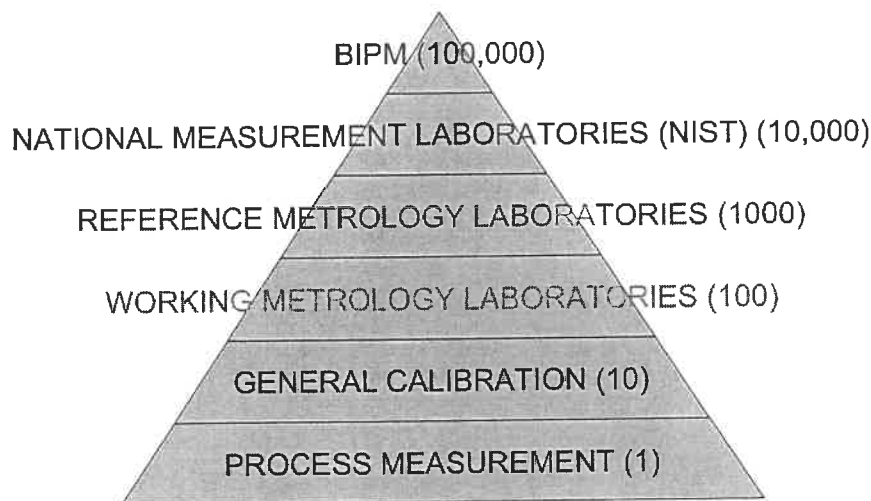


Figure 1: The 10:1 TAR Ratio and Calibration Traceability

As one moves up the calibration traceability pyramid, it becomes apparent that standards capable of achieving accuracy levels required are just not existent. The term accuracy is also misunderstood in defining this requirement. Sometimes the resolution of the instrument is the sole criteria defined in the accuracy of the instrument with no regard to other instrument characteristics that go into defining the accuracy specification.

To overcome this difficulty of achieving 10:1 TAR, the 4:1 TAR was introduced and judged acceptable. However, the 4:1 TAR becomes difficult to achieve as one moves up the traceability pyramid illustrated in Figure 2.

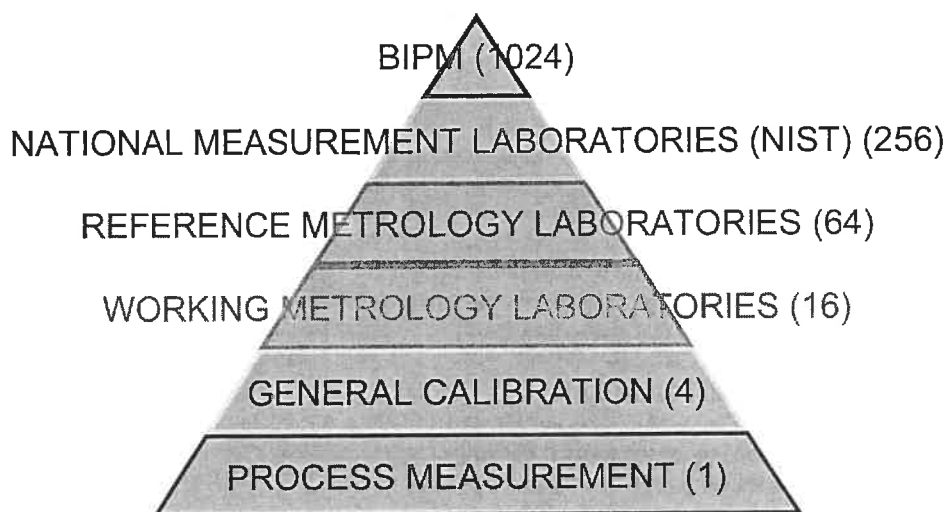


Figure 2: The 4:1 TAR Ratio and Calibration Traceability

Based on the definition of traceability, the uncertainties are cumulative. Thus the calibration traceability pyramid is now more realistic in making the measurement claims through use of estimated uncertainties. Used properly, the traceable calibration with measurement uncertainty is a more credible way to report the calibration data than making TAR claims.

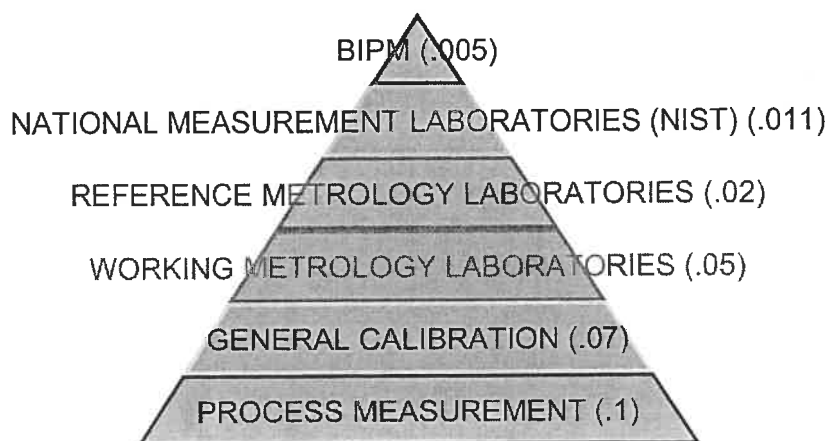


Figure 3: Calibration Traceability Pyramid

4.0 MEASUREMENT UNCERTAINTY

Other factors influence in making a measurement. These factors influence the overall measurement and complement in estimating the measurement uncertainty besides the use of manufacturer's stated specifications.

According to the GUM, the process of estimating measurement Uncertainty can be broken down in the following simplified steps. The author has identified several common errors observed while working with both the calibration laboratories and their customers.

1. Identify the uncertainties in the measurement process.

The factors contributing to the uncertainty can come from a variety of sources. It is very common for the industry to ignore many factors affecting the measurement process. Environmental considerations, operator to operator or operator to equipment interactions are often ignored. Long term stability data is available but rarely included in the uncertainty budgets. Many laboratories do not put enough effort in this exercise.

2. Classify type of uncertainty (A or B). Quantify (evaluate and calculate) individual uncertainty by various methods.

Type A evaluation method: The method of evaluation of uncertainty of measurement by the statistical analysis of series of observations.

Repeatability and Reproducibility analysis can be easily determined with computer software tools yet this information is not included in the uncertainty budget considerations. Analysis techniques such as Statistical Process Control (SPC), Analysis of Variance (ANOVA) and Design of Experiments (DOE) are not utilized as often.

Type B evaluation method: The method of evaluation of uncertainty of measurement by means other than the statistical analysis of series of observations.

While more estimates of Type B uncertainty components are included in the budget, Type B estimates of uncertainty are not sufficiently analyzed. When included in the uncertainty budgets, standard uncertainty is not derived using the familiar correction factors depending on rectangular, triangular or U-shaped distributions.

3. Document in an uncertainty budget.

Uncertainty budgets do not contain enough information to determine how the estimates are derived. Once the budgets are developed, they get filed and are not reviewed as often when the calibration process changes or new equipment is added. It is difficult to second guess how the information was derived if sufficient information (including raw data) is not included in budget.

4. Combine uncertainty (Root Sum Square (RSS) method).

$$u_{c_a} = \sqrt{u_{c_{a1}}^2 + u_{c_{a2}}^2 + \dots} \quad \text{Combine Type A uncertainties}$$

$$u_{c_b} = \sqrt{u_{c_{b1}}^2 + u_{c_{b2}}^2 + \dots} \quad \text{Combine Type B uncertainties}$$

$$u_c = \sqrt{u_{c_a}^2 + u_{c_b}^2} \quad \text{Combine all Type A and Type B uncertainties}$$

The root sum square method is used for combining all the uncertainty components. No determination is made for correlated components of uncertainty. In most cases, this does not penalize the uncertainty budget as much (remember: it is an estimate), in some critical cases, it does misrepresent data to make a difference.

5. **Assign appropriate k factor multiplier to combined uncertainty to report expanded uncertainty.**

$$U = k \cdot u_c \quad \text{The GUM specifies } k = 2 \text{ or } 95\% \text{ confidence interval.}$$

Although, this is a straight multiplier, industry in general does not understand the use of confidence intervals and degrees of freedom. Sometimes, misstating uncertainty at $k=1$ makes the numbers look good on the scope of accreditation. Various numbers games are played using uncertainty numbers in what is called the "Scope Wars" among commercial calibration laboratories. It is the responsibility of the accrediting body to ensure that a laboratory's scope of accreditation is stated accurately. However, there is misrepresentation of data in the scope of accreditation.

5.0 **OTHER INDUSTRY MISAPPLICATION OF MEASUREMENT UNCERTAINTY ESTIMATES**

From the calibration service user's point of view, it is easier to state that the measurement equipment is calibrated with its "badge of honor" (namely the calibration sticker) than understand the measurement uncertainty and its proper application. The uncertainty stemming from the industry in not knowing how measurement uncertainty estimates are applied in their daily application results in various misapplications of measurement uncertainty. They are listed below.

1. Equipment with larger uncertainty than the required tolerance is used to make measurements.
2. Equipment with very small uncertainty is used where required tolerance is large.
3. Judgment is based on a single measurement on a critical parameter.
4. Lack of compensation for environment when measurements are made.
5. More significant digits on measurement equipment mean more accuracy.
6. Making a measurement with equipment with less resolution and increasing the precision (i.e. more decimal places) by manipulating the numbers in a spreadsheet.
7. Claiming very low uncertainties on measuring equipment incapable of resolving such low values.
8. Lack of a measurement assurance program to ensure that the estimated measurement uncertainty has not changed and assuring confidence in the measurement process over a period of time.

6.0 **SUMMARY**

This paper outlined some misapplications and errors sources contributing to the uncertainty of the (pun intended) application of measurement uncertainty by the both the industry and calibration laboratories alike. The GUM is a great reference document for the experienced metrologist who is also a statistician. However, it is not an easy document to read for a common practitioner or a user of calibration services. Guide supplements to GUM for the user of calibration services (Part A) and practitioner (Part B) should help resolve this confusion. As we move toward a very global economy, the one common force that drives the world commerce will be the measurement process. Quantifying and understanding the uncertainties associated with the measurement become even more important as we apply it to other non-traditional uses besides metrology.

7.0 REFERENCES

1. Group, Z540-2, 1997, *U. S. GUIDE TO EXPRESSION OF UNCERTAINTY IN MEASUREMENT*, Boulder, ANSI/NCSL International
2. ISO 1993. International vocabulary of basic and general terms in Metrology (VIM), Geneva: ISO
3. ANSI/ISO 17025-1999, American National Standard-General requirements for the competence of testing and calibration laboratories. Milwaukee: ASQ Quality Press
4. Bucher, Jay L. 2004. *The Metrology Handbook*. Measurement Quality Division. Milwaukee: ASQ Quality Press

Industrial Applications of Uncertainty Analysis

Jonathan W. Arenberg
Northrop Grumman Space Technology
One Space Park Drive
Redondo Beach, CA 90278

Abstract:

Uncertainty or tolerance analysis is a powerful technique for addressing many problems facing the practicing engineer. This paper reviews several of the common uses of this tool and introduces some novel applications.

The paper begins with derivation of "Law of Error Propagation" and follows the developments presented earlier.¹ In this development, explicit forms for the outcome of a process and its statistics are developed.

These results are then applied to a number of common industrial applications. The applications include, error budgeting and requirements derivation. Included in the discussion of the derivations of requirements is a discussion of the need to include margin and contingency explicitly into the analysis.

Next, an expression for the repeatability of a process is derived. This analysis is then used to identify the general conditions that lead to a repeatable process.

Combining cost data or modeling, error budgets are extended into tools that can be used in Cost as an Independent Variable (CAIV) modeling and analysis. The combined model can then be used to determine cost as function of tolerance and the marginal cost of repeatability.

Introduction

This paper derives the Law Of Error Propagation and explores its uses in determining the outcome and distribution of a process in the presence of variations. The explicit purpose of this paper is tutorial in nature, specifically to fill a void in the existing literature on the origins of this important result. In starting from first principles, significant insight can be gleaned on the wide applicability and utility of this powerful Law.

The derivation begins with a general mathematical model of a measurement or process. The sensitivities of the expected outcome of the process to variances in the individual parameters of the problem are derived. The analysis continues and gives expressions for the expectation value (most likely outcome) of the process, μ , in the presence of the variances. A second product of this derivation is the variance of the distribution about μ , generally denoted, σ^2 . The derivation

¹ J.W. Arenberg, "On the Origins of the Law of Error Propagation and its Uses ", American Society for Precision Engineering Summer Topical Meeting on Tolerancing, Charlotte, N.C., July 2002.

clearly identifies the assumptions leading to the formula for error propagation, and under what conditions the traditional rss formula is applicable. It will be shown that the traditional rss expression is a special case of the more general Law of Error Propagation, where all parameters of the problem are not independent. The paper will also clearly identifies which statistic is to be combined in an rss fashion.

Outcome of a Process

This section derives the expression for the mean and variance of a process in the presence of variations about an ideal or desired set of parameters. These outcomes, mainly the expression for the variance are well known as the Law of Error Propagation.

Consider a process or procedure of many variables that has an outcome z . The outcome can be the result of a manufacturing process, measurement or even a high level system performance parameter. The outcome, z , is called the measurand, and can be represented as the mathematical process, f , operating on the parameters or variables of the problem denoted, $w_1, w_2 \dots w_N$ and is written

$$z = f(w_1, w_2, \dots, w_N) \quad (1)$$

Let the optimum, desired or "true" value of z be denoted, μ_z the set of w_i that give the value of μ_z , are denoted μ_i . Thus, the optimum process is written

$$\mu_z = f(\mu_1, \mu_2, \dots, \mu_N). \quad (2)$$

Each w_i has a set of likely values according to its own probability distribution, $p_i(w_i)$. If the probability densities of the w_i about μ_i are known, it is possible to calculate directly the distribution of likely outcomes of z . Assuming that the probability distributions are functions of only one w_i and are independent, the probability density of an outcome z is given by

$$p(z) = f(w_1, w_2, \dots, w_N) \prod_{i=1}^N p(w_i) \quad (3)$$

Most often (3) is evaluated using Monte Carlo techniques. Armed with the knowledge of the probability density, all of the moments of z can be calculated directly. A frequently used process is to fit the outcome of the Monte Carlo calculation (usually a histogram) To a known distribution for further analysis.²

A useful application is estimating the probability of meeting a specific requirement, denoted R (where higher is a better performance). This probability is written

² A recent example is given in "An intermodule alignment budget for the Generation X telescope", Jonathan Arenberg and Christopher Lau, in SPIE 5488 (2004). To be published.

$$\Pr(z \geq R) = \int_R^{\infty} p(z) dz \quad (4)$$

If $p(z)$ has been parameterized in terms of μ_i and σ_i then the probability of meeting a given performance can be expressed in terms of the design point and tolerance set.

In some cases, it is sufficient to only calculate, σ_z . In these situations, small deviations from the μ_i and their impact on μ_z are calculated from a Taylor Series expansion about the μ_i which is

$$z = \mu_z + \sum_{i=1}^N \frac{\partial f}{\partial w_i} (w_i - \mu_i) + O((w_i - \mu_i)^2) \quad (5)$$

Ignoring terms higher than first order in $(w_i - \mu_i)$ gives

$$z = \mu_z + \sum_{i=1}^N \frac{\partial f}{\partial w_i} (w_i - \mu_i) \quad (6)$$

Equation (6) is the first major result of this derivation. It gives the value of the measurand in the case where the w_i are not equal to the μ_i . Next, μ_z is subtracted from both sides of (6) to give

$$z - \mu_z = \sum_{i=1}^N \frac{\partial f}{\partial w_i} (w_i - \mu_i) \quad (7)$$

The term on the left hand side is the deviation from the ideal value, μ_z . Squaring (7) gives

$$(z - \mu_z)^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial w_i} \right)^2 (w_i - \mu_i)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial w_i} \right) \left(\frac{\partial f}{\partial w_j} \right) (w_i - \mu_i) (w_j - \mu_j) \quad (8)$$

Next, the expectation value, $E(x)$ ³, is taken of both sides of (8)

$$E[(z - \mu_z)^2] = E \left[\sum_{i=1}^N \left(\frac{\partial f}{\partial w_i} \right)^2 (w_i - \mu_i)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial w_i} \right) \left(\frac{\partial f}{\partial w_j} \right) (w_i - \mu_i) (w_j - \mu_j) \right] \quad (9)$$

Since the expectation value of a sum is the sum of the expectation values, the right hand side of (9) is written

³ The expectation value $E(x)$ is given by the expression $E(x) = \int_{-\infty}^{\infty} xp(x) dx$, where $p(x)$ is the probability distribution of x .

$$\sigma_z^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial w_i} \right)^2 E \left[(w_i - \mu_i)^2 \right] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial w_i} \right) \left(\frac{\partial f}{\partial w_j} \right) E \left[(w_i - \mu_i)(w_j - \mu_j) \right]. \quad (10)$$

The change in the left hand side of (10) is made since the variance of z , is the expectation value of, $z - \mu_z$, namely, $\sigma_z^2 = E(z - \mu_z)^2$. Equation (10) reduces to

$$\sigma_z^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial w_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial w_i} \right) \left(\frac{\partial f}{\partial w_j} \right) \sigma_{ij} \quad (11)$$

σ_{ij} is called the covariance of i and j . The covariance is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (12)$$

where ρ_{ij} is the correlation coefficient. So,

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}. \quad (13)$$

Substitution of (13) into (11) gives

$$\sigma_z^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial w_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial w_i} \right) \left(\frac{\partial f}{\partial w_j} \right) \rho_{ij} \sigma_i \sigma_j. \quad (14)$$

Equation (14) is the general expression of the variance in the measurand. No assumptions about the nature of the $p_i(w_i)$ were made and so the result, (14) is quite general and applicable.

For parameters that are independent, the covariance is zero, and (14) becomes

$$\sigma_z^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial w_i} \right)^2 \sigma_i^2. \quad (15)$$

which is the well known result and the goal of this derivation. Also note that the σ_i^2 are the variances for each variable i . From this derivation is clear that the quantities that are rss'ed are the σ_i and not some other statistic.

Error Budgeting

A chief industrial application of this class of analysis is the error (tolerance or performance budget). This is a device to capture and communicate the design point and more commonly the

allowed deviations from the design point. The error budget is the set of deviations from $\bar{\mu}$ that still satisfy requirements. These errors are most usually and easily expressed as the standard deviations, σ_i , around the μ_i . The process of adjusting difficulty and cost among the σ_i is referred to as error budgeting.

Discussion of the error budget will be extended to include a procedure to determine adequate margin and contingency so that a specific yield (probability of success) can be attained.

Repeatability

The repeatability of a process is the probability that a second outcome agrees to within ε of the first. If the this probability is denoted, $R(\varepsilon)$ and the two outcomes are Λ_1 and Λ_2 , the $R(\varepsilon)$ may be written

$$R(\varepsilon) = \Pr(|\Lambda_1 - \Lambda_2| \leq \varepsilon) = \Pr(\Lambda_1) \Pr(\Lambda_2 \leq |\Lambda_1 - \varepsilon|) \quad \forall \quad \Lambda_1. \quad (16)$$

The probability density of Λ is $p(z)$ namely, so (16) becomes

$$R(\varepsilon) = \int_0^{\infty} p(\Lambda_1) dz_1 \cdot \Pr(\Lambda_2 \leq |\Lambda_1 - \varepsilon|) \quad (17)$$

which is

$$R(\varepsilon) = \int_0^{\infty} p(\Lambda_1) d\Lambda_1 \left\{ \int_{\Lambda_1 - \varepsilon}^{\Lambda_1 + \varepsilon} p(\Lambda) d\Lambda \right\}. \quad (18)$$

Equation (18) can be written in equivalent form in terms of the cumulative distribution of the first order statistic $\Psi(x_{(1)})$

$$R(\varepsilon) = \int_0^{\infty} p(\Lambda_1) \{P(\Lambda_1 + \varepsilon) - P(\Lambda_1 - \varepsilon)\} d\Lambda_1, \quad (19)$$

where

$$P(\Lambda) = \int_0^{\Lambda} \psi(\Lambda') d\Lambda'. \quad (20)$$

The value of $R(\varepsilon)$ will be largest when the term in the braces is large, either via large value of ε , or a tight distribution (steep cumulative function).

Other Applications

As shown above in (4), the probability of meeting a given requirement can be determined for a specific process, design point and tolerance set. If the $\Pr(z \geq R)$ is denoted Q (Q for quality) and

cost relationships or models exist for the m_i and s_i and expression for the cost, C of having a given probability of equaling or exceeding a specific outcome can be formulated, $C(R, \bar{\mu}, \bar{\sigma})$. Q can be plotted against C for various design points and tolerance sets. This analysis allows for direct examination of cost versus quality and what elements in the design and tolerance space are most influential.

Conclusion

This paper has presented several common industrial applications of uncertainty analysis. Extensions to these basic techniques will be discussed in the oral presentation. It is hoped that this paper will contribute to the propagation of these methods and associated formalism.

Overview of Uncertainty Analyses for Gauge Block Calibration

Jennifer E. Decker, René Schödel*, Gerhard Bönsch*
Institute for National Measurement Standards (INMS),
National Research Council Canada (NRC),
Ottawa, Canada, K1A 0R6
*Physikalisch-Technische Bundesanstalt (PTB),
Bundesallee 100, 38116 Braunschweig, Germany

Gauge blocks continue to be essential reference artifacts in manufacturing because of their simple geometry, but also because they can provide very high accuracy for a reasonable cost. The calibration technique of optical interferometry can directly link gauge block central length to the distance light travels in vacuum in $1/299792458$ of a second, the definition of the International System of Units (SI) of length, the metre [1]. Gauge blocks calibrated by interferometry are then used in turn as reference standards to calibrate gauge blocks by mechanical comparison techniques, subsequently propagating the definition of the metre along the chain of calibrations, each with a specified measurement uncertainty. Even though these two calibrations involve the same artifact, dominant components in the evaluations of measurement uncertainty for these two techniques are fundamentally different. This is due to the fact that the method of calibration for interferometry involves measuring light waves in air and absolute gauge block temperature, whereas mechanical comparison involves tactile probing and the important temperature measurement is actually the temperature difference between reference standard and client gauge blocks. Detailed measurement uncertainty evaluations in accordance with the ISO Guide to the Expression of Uncertainty in Measurement (GUM) [2] for calibration of gauge blocks by techniques of interferometry and mechanical comparison can be found in the literature (see below). Some of the dominant components of these uncertainty budgets are highlighted for discussion.

Uncertainty Evaluations for Gauge Block Calibration – Some Good Examples: Gauge block calibration is one of the mainstay's of many calibration and measurement labs; moreover the geometrical simplicity of gauge blocks makes them an ideal subject for instructive documentation on uncertainty evaluation. Therefore there are several interpretations of gauge block measurement uncertainty evaluations in the literature. A general uncertainty evaluation should be applied thoughtfully to an individual measurement set-up, carefully making sure the components of the budget correctly represent the particularities of the individual measurement.

Hariharan [3] gives a brief but accurate description of what an interferometer is and how the method of exact fractions for gauge block length measurement by optical interferometry works. Several versions of uncertainty budgets for gauge block calibration by interferometry are published in the literature [4, 5, 6, 7], each corresponding to slightly different optical set-ups and routine practices (see below for mention of contributions for refractive index of air).

For gauge block calibration by dual probe mechanical comparison, an uncertainty budget for reference and client gauge blocks made of the same material can be found in the GUM [8], the summary table from which can also be found in the NIST reference on uncertainty evaluation [9]. Decker and Pekelsky [10] provides detailed worked example along this same formalism, including detailed consideration of the higher-order uncertainty contributions. In this model the higher-order components manifest almost all the thermal contributions. The different versions of uncertainty budgets differ in the way that the model equation is manipulated before the GUM guidelines are applied. In the example of ISO/TS 14253-2:1999(E) *Guide to the estimation of uncertainty in GPS measurement in calibration of measuring equipment and in product verification* [11] which outlines the Procedure for Uncertainty of Measurement Management (PUMA) the equations describing thermal contributions are completely general, all contributions are treated in first order (drift testing is helpful for this as described below) and equations describing the correlation contributions are provided. Yet another model describing an averaging or substitution-method

model equation is adopted in [12]. Another detailed example employing a very similar model equation can be found (for free) in the European co-operation for Accreditation publication, which also gives some useful insight on measurement uncertainty evaluation [13].

A word about the PUMA method: it applies GUM guidelines in a series of iterations where a very conservative analysis is the starting point. Based on whether or not the resulting total expanded uncertainty is satisfactory for the application, each component in the uncertainty budget is then re-visited to see if more information or measurements would yield an uncertainty that is desired. The premise is that it is acceptable if the uncertainty is greatly overestimated, so long as its value meets requirements. This method of analysis is convenient in applications where uncertainty evaluations are required primarily for quality documentation obligations. However, in applying this procedure one should already have a feeling as to the dominant uncertainty contributions to the measurement. As a note of caution – sometimes experimental evidence uncovers components that unexpectedly turn out to be large.

Preparing a measurement uncertainty evaluation has the potential for being a thorough learning tool. For example, the relative importance of measurement influences become immediately apparent. A straightforward application of GUM shows immediately where improvements to the calibration set-up would have the most value. For example, in our experience, uncertainties associated with the measurement of air pressure made surprisingly large contributions through evaluation of the refractive index of air by empirical equations. It was large enough to be a significant contributor to the overall gauge block uncertainty budget, but was easily remedied by purchasing and calibrating a new, more accurate, pressure sensor. Refractive index of air is one of the largest corrections to any length measurement made by interferometric methods, and therefore the uncertainty contributions are also very important since they can dominate the uncertainty budget. Ideas on how the measurement of air temperature, air pressure and relative humidity influences the application of empirical equations or refractometer methods can be found in the discussion of this topic in the literature [14, 15, 16, 17] (and references therein). A couple of articles of particular relevance for shop-floor or Cal-lab applications are [18, 19].

Uncertainties associated with end effects vs. length-dependent influences can make the same uncertainty evaluation look completely different simply depending on the nominal length of the gauge block. End effects dominate the uncertainty of short gauge blocks (interferometry: fringe fraction measurement, wringing, phase change on reflection, optical wavefront errors, gauge block geometry; mechanical comparison: reference gauge calibration, measured length difference, correction for elastic deformation between reference and client gauge), whereas the length dependent influences dominate the uncertainty of long gauge blocks (interferometry: obliquity, temperature, refractive index of air; mechanical comparison: temperature). For example, refractometer techniques offer a lower uncertainty alternative for long gauge block lengths (125 mm to 1 m nominal length), whereas for gauge blocks shorter than about 100 mm the end effect components dominate the total uncertainty budget and installing a refractometer in the set-up would not offer very much advantage. Similarly, gauges with almost-zero value of thermal expansion coefficient will change the relative size of components in the uncertainty budget, which means that they can be a useful diagnostic tool.

Measurement uncertainty evaluation for mechanical comparison where reference and client gauge blocks are dissimilar materials can be dominated by the first order temperature contributions which now become very large. Moreover, the uncertainty contribution for elastic deformation correction [20, 21, 22] can also make a significant contribution. A detailed treatment of measurement uncertainty for dissimilar materials can be found in [23, 24]. Further detailed considerations regarding uncertainty evaluation for corrections of elastic deformation or thermal environment have been proposed for the next updated version of the standard ASME B89.1.2 *Calibration of Gage Blocks by Contact Comparison Methods*.

Phase Stepping or Phase Shifting Applications: Phase step techniques are becoming increasingly popular in many length- and/or form-analysing instruments. The output of these interferometer instruments is a phase map of the optical wavefront, or the interferometer image field of view. The measurand for length applications is usually a phase difference between different areas of the image, which is often converted to a fractional order of an interference fringe for further analysis by the method of exact fractions. These instruments offer the advantageous potential for very precise measurement of interference fringes, however the disadvantage is that the systematic component of measurement uncertainty associated with the phase readings can be dependent on the actual phase value measured. In most commercial instrumentation the errors that contribute to these systematic components such as phase shifter errors, or camera nonlinearities, have been characterized and minimized by the instrument

manufacturer. However for confidence in claims of better than 0.01 fringe measurement uncertainty these errors should be characterized experimentally. These considerations have been discussed in [25], along with correlation between phase measurements from two parts of the same image that used for refractivity and length evaluation, respectively. It turns out that these correlations can be small if the systematic components are also small, but that they are length dependent, reflecting the length-dependent influence of the refractive index of air (larger error in the phase measurement has larger impact on evaluated refractivity). Also of concern for precision instrumentation applying phase stepping analysis is the accurate and repeatable location of the reference point on a gauge in an image, and the assignment of an associated uncertainty to the length measurement. For high-precision applications such errors can be significantly large and the uncertainties associated with these measurements are best evaluated experimentally [26].

Measuring a length difference relative to a reference can have smaller uncertainty than measuring a similar absolute length. In measuring the length difference between two similar gauge blocks when the thermal influences are well controlled, the difference measurement adds only a small component to the uncertainty of the reference standard (for the same gauge block materials). Thermal influences are discussed in detail in the ISO/TR 16015:2003(E) *Systematic errors and contributions to measurement uncertainty of length measurement due to thermal influences* standard [27]. This document describes techniques for carrying out and analyzing results from empirical drift testing to result in realistic length-equivalent corrections and estimations of measurement uncertainty. The drift test technique integrates all influences and therefore the mathematical modeling and analysis need not be so sophisticated in order to completely capture all measurement influences. Given the unique high-precision and digital image processing characteristics offered by phase step interferometry techniques, differential measurement of gauge blocks by optical interferometry is being considered as a method to further reduce uncertainty in gauge block calibration.

The authors thank Jim Pekelsky and Kos Doytchinov of NRC, Canada for fruitful discussion. The authors gratefully acknowledge funding from the Alexander von Humboldt Foundation, Germany.

References

- [1] T. J. Quinn, "Practical realisation of the definition of the metre (2001)," *Metrologia* **40**, pp. 103–133, 2003.
- [2] ISO *Guide to the Expression of Uncertainty in Measurement* (GUM), International Organization for Standardization (ISO), Switzerland, 1993.
- [3] P. Hariharan, *Basics of Interferometry*, (Academic Press, Inc., San Diego), 213 pages, 1992.
- [4] K. G. Birch, "Uncertainties in the Measurement of Gauge Blocks by Interferometry," National Physical Laboratory (NPL), Teddington, Middlesex, UK, Report MOM 29, May 1979.
- [5] F. Bayer-Helms, "Gauge blocks: concepts, standardization, measurement," Physikalisch-Technische Bundesanstalt, PTB-Bericht Opt-28e, Braunschweig, Germany, February 1989. (ISBN 3-88 314-864-4)
- [6] H. Darnedde, "High-precision Calibration of Long Gauge Blocks Using the Vacuum Wavelength Comparator," *Metrologia* **29**, pp. 349–359, 1992.
- [7] J. E. Decker and J. R. Pekelsky, "Uncertainty evaluation for the measurement of gauge blocks by optical interferometry," *Metrologia* **34**, pp. 479–493, 1997.
- [8] ISO GUM, op. cit. §H.1.
- [9] Barry N. Taylor, C. E. Kuyatt, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, US Department of Commerce, NIST Technical Note 1297, 20 pages, 1994.
- [10] J. E. Decker and J. R. Pekelsky, "Uncertainty of Gauge Block Calibration by Mechanical Comparison: A Worked Example, Gauges of Like Material," Presented at the National Conference of Standards Laboratories (NCSL) Canadian Region Spring Meeting, Ottawa, Canada, 16 May 1996. NRC Document No. 39998. http://inms-ienm.nrc-cnrc.gc.ca/research_and_development/orange.pdf

- [11] ISO/TS 14253-2:1999(E) *Geometrical Product Specifications (GPS) — Inspection by measurement of work-pieces and measuring equipment — Part 2: Guide to the estimation of uncertainty in GPS measurement, in calibration of measuring equipment and in product verification*, International Organization for Standardization (ISO), Switzerland, 1999.
- [12] G. Bönsch, W. Kessel, "Kalibrierung von Parallelendmassen aus gleichem Werkstoff durch Unterschiedsmessung (Substitutions-Messmethode)," PTB-MU-L, 19 pages, 23 April 2002.
- [13] *Expression of the Uncertainty of Measurement in Calibration*, European co-operation for Accreditation, Document EA-4/02, 79 pages, December 1999. (available in Adobe .pdf format: <http://www.european-accreditation.org>)
- [14] P. E. Ciddor, "The refractive index of air: new equations for the visible and near infrared", *Applied Optics* **35**(9) 1566–1573, 1996.
- [15] K. P. Birch and M. J. Downs, "Correction to the updated Edlén equation for the refractive index of air", *Metrologia* **31** 315–316, 1994.
- [16] G. Bönsch and E. Potulski, "Measurement of the refractive index of air and comparison with modified Edlén's formulae," *Metrologia* **35**, pp. 133–139, 1998.
- [17] J. E. Decker, R. Schödel, G. Bönsch, "Next-Generation Kösters Interferometer," *Proc. of SPIE Recent Developments in Traceable Dimensional Measurements*, Editors J. E. Decker, N. Brown, San Diego, USA, SPIE Vol. **5190**, 14–23, 2003.
- [18] K. P. Birch, F. Reinboth, R. E. Ward, G. Wilkening, "The Effect of Variations in the Refractive Index of Industrial Air upon the Uncertainty of Precision Length Measurement," *Metrologia* **30**, pp. 7–14, 1993.
- [19] W. Tyler Estler, "High-accuracy displacement interferometry in air," *Applied Optics* **24**, pp. 808–815, 1985.
- [20] M. J. Puttock, E. G. Thwaite, "Elastic Compression of Spheres and Cylinders at Point and Line Contact," Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia, Technical Paper No. 25, 65 pages, 1969.
- [21] NIST Engineering Metrology Toolbox, <http://patapsc.nist.gov/mel/div821/>
- [22] *The Gage Block Handbook*, Edited by Ted Doiron and John S. Beers, NIST Monograph 180, (US Department of Commerce), 142 pages, 1995.
- [23] J. E. Decker and J. R. Pekelsky, "Uncertainty of Gauge Block Calibration by Mechanical Comparison: A Worked Example, Case 2: Dissimilar Materials," *Proc. of SPIE Recent Developments in Optical Gauge Block Metrology*, Editors J. E. Decker, N. Brown, San Diego, USA, SPIE Vol. **3447**, pp. 225 - 246, 1998.
- [24] J. E. Decker, A. Ulrich, A. Lapointe, M. Viliesid, J. R. Pekelsky, "Two-part study towards lowest uncertainty calibration of ceramic gauge blocks: interferometry and mechanical comparison techniques," *Proc. of SPIE Recent Developments in Traceable Dimensional Measurements*, Editors J. E. Decker, N. Brown, Munich, Germany, SPIE Vol. **4401**, pp. 23 - 32, 2001.
- [25] J. E. Decker, R. Schödel, G. Bönsch, "Considerations for the evaluation of measurement uncertainty in interferometric gauge block calibration applying methods of phase step interferometry," *Metrologia* **41**, pp. L11–L17, 2004.
- [26] R. Schödel, J. E. Decker, "Methods to recognize the sample position for most precise interferometric length measurements," *Proc. of SPIE Conference Interferometry*, Denver, Colorado, August 2004.
- [27] ISO/TR 16015:2003(E) *Geometrical Product Specifications (GPS) — Systematic errors and contributions to measurement uncertainty of length measurement due to thermal influences*, International Organization for Standardization (ISO), Switzerland, 2003.

IN-SITU CALIBRATION OF A REDUNDANT MEASUREMENT SYSTEM FOR MANIPULATOR POSITIONING

Piotr J. Meyer

Philips Center for Industrial Technology (CFT), Lynnfield, MA, U.S.A.

Prof. Dr. Jan van Eijk

Philips Center for Industrial Technology (CFT), Eindhoven, The Netherlands

Keywords: Manipulator Positioning, Measurement System, Displacement Sensor, Metrology Model, Redundancy, In-Situ Calibration, Machine Errors.

1. Introduction

Achieving high positioning accuracy in a precision manipulator usually relies on calibration of its metrology system, i.e. the determination of reproducible geometry errors as represented by parameters in a metrology model, which is then used for position control.

The issue of metrology calibration in the presence of external sensors, which can provide reliable information on the true spatial position of the manipulator's end-effector, is well understood. A problem arises, however, in the absence of such external sensors, when the manipulator's internal sensors, all of them encumbered with unknown geometry errors, provide the entire position information.

Many high-performance manipulators feature a redundant metrology system. In such a system more spatial measurements are available than the number of degrees of freedom. This paper shows how that redundancy can be utilised to calibrate the manipulator in situ, i.e. without the use of external measurement systems.

2. Coordinate Transformations for Redundant Metrology

Our method requires introducing the inverse and forward coordinate transformations for the measurement system in a formal manner. Firstly, the **inverse coordinate transformation** maps the 6 DoF spatial position $\mathbf{p}$ of the end-effector into the sensor readings $\mathbf{q}$:

$$\mathbf{q} = \mathbf{f}(\mathbf{p}, \mathbf{g}) \quad (1)$$

Such a transformation essentially constitutes the metrology model and is uniquely defined. The unknown geometry errors $\mathbf{g}$ are included as parameters in (1), either through analytical modelling of the manipulator geometry or through arbitrary expressions, such as polynomials. Both linear

and non-linear treatment of the parameters is possible. It is the purpose of calibration to determine the geometry parameters $\mathbf{g}$.

Secondly, the **forward coordinate transformation** maps the measurement values $\mathbf{q}$ into the 6 DoF spatial position $\mathbf{p}$ of the end-effector:

$$\mathbf{p} = \mathbf{f}^+(\mathbf{q}, \mathbf{g}) \quad (2)$$

This is the transformation which allows us to control the manipulator, provided we know the geometry parameters $\mathbf{g}$. It is not uniquely defined due to redundancy and has to be based on an arbitrary decision that its result always satisfy some of the measurements, and minimise the differences over the remaining measurements in the least-square sense. In particular, the simplest assumption would be a uniform least-square fit over all the sensors.

For consistency, we must demand that transforming a given 6 DoF position into sensors values and then transforming the latter back to a 6 DoF position result in the original position. The converse is not generally true, though, due to the fact that $\mathbf{q}$ is redundant and does not have to be consistent with our model. Formally:

$$\begin{aligned} \mathbf{f}^+(\mathbf{f}(\mathbf{p}, \mathbf{g}), \mathbf{g}) &= \mathbf{p} \\ \mathbf{f}(\mathbf{f}^+(\mathbf{q}, \mathbf{g}), \mathbf{g}) &\neq \mathbf{q} \end{aligned} \quad (3)$$

In this sense the forward transformation is not a full inverse of the inverse transformation, but a “pseudo-inverse”, by analogy with linear algebra.

3. Computation of the Forward Transformation

Computing the inverse transformation does not pose any problems, however the forward transformation is not available in the closed form and has to be computed with the use of numerical methods. If, as previously suggested, we aim at the best fit to our redundant measurements, we seek the minimum of the fit error E (here, we omit the parameters $\mathbf{g}$ for brevity):

$$\begin{aligned} E &= \|\mathbf{f}(\mathbf{p}) - \mathbf{q}\|_2 = (\mathbf{f}(\mathbf{p}) - \mathbf{q})^T (\mathbf{f}(\mathbf{p}) - \mathbf{q}) \\ \frac{\partial E}{\partial \mathbf{p}} &= 2 \left(\frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right)^T (\mathbf{f}(\mathbf{p}) - \mathbf{q}) = \mathbf{0} \end{aligned} \quad (4)$$

Denoting the Jacobian of the inverse transformation by $\mathbf{A}$:

$$\mathbf{A} = \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \quad (5)$$

we can write the equation, whose unique solution is the stage position $\mathbf{p}$, as follows:

$$\mathbf{F}(\mathbf{p}) = \mathbf{A}^T (\mathbf{f}(\mathbf{p}) - \mathbf{q}) = \mathbf{0} \quad (6)$$

A Newton iteration follows automatically:

$$\begin{aligned} \mathbf{p}_{n+1} &= \mathbf{p}_n - \left(\frac{\partial \mathbf{F}}{\partial \mathbf{p}} \right)^{-1} \mathbf{F}(\mathbf{p}_n) = \\ &= \mathbf{p}_n - (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T (\mathbf{f}(\mathbf{p}_n) - \mathbf{q}) = \\ &= \mathbf{p}_n - \mathbf{A}^+ (\mathbf{f}(\mathbf{p}_n) - \mathbf{q}) \end{aligned} \quad (7)$$

Denoting the matrix which appears in front of the error term, in this case a pseudo-inverse of $\mathbf{A}$, by $\mathbf{G}$:

$$\mathbf{G} = \mathbf{A}^+ \quad (8)$$

we observe that, at the limit of the Newton iteration, $\mathbf{G}$ becomes the Jacobian of the forward transformation:

$$\begin{aligned} \mathbf{p}_n - \mathbf{p}_{n+1} &= \mathbf{G} (\mathbf{f}(\mathbf{p}) - \mathbf{q}) = \mathbf{G} (\mathbf{f}(\mathbf{f}^+(\mathbf{q})) - \mathbf{q}) = \mathbf{0} \\ \mathbf{G} \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} - \mathbf{G} &= \mathbf{0} \Rightarrow \mathbf{G} \mathbf{A} \frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} = \mathbf{G} \Rightarrow \frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} = \mathbf{G} \end{aligned} \quad (9)$$

If our arbitrary choice of the forward transformation is different than the uniform least-square fit, the form of the matrix $\mathbf{G}$ will be more complex. In fact any matrix $\mathbf{G}$ which satisfies

$$\mathbf{G} \mathbf{A} = \mathbf{I} \quad (10)$$

is admissible as the Jacobian of a valid forward transformation. One can convince oneself that this is a necessary condition by differentiating the first line of (3) over $\mathbf{p}$.

4. Calibration Equation

Once a manipulator moves through its workspace, the only reliable information at our disposal is the actual sensor readings, which we will denote by $\mathbf{q}_1$. The manipulator, in reality, corresponds to a certain vector $\mathbf{g}_1$ of geometry parameters, which is unknown to us. Instead, we use another vector $\mathbf{g}_0$ of uncalibrated parameters, e.g. their nominal values. Let us apply the composition of the inverse and forward coordinate transformation to the actual sensor readings, assuming in both operations the incorrect parameters $\mathbf{g}_0$:

$$\mathbf{q}_0 = \mathbf{f}(\mathbf{f}^+(\mathbf{q}_1, \mathbf{g}_0), \mathbf{g}_0) \neq \mathbf{q}_1 \quad (11)$$

We shall call the resulting $\mathbf{q}_0$ **anticipated sensor readings**. If we had the knowledge of the correct parameters $\mathbf{g}_1$ instead of $\mathbf{g}_0$, the anticipated sensor readings would be identical with the actual ones. This suggests that the difference between the actual and anticipated readings is related to the mismatch between the real manipulator ($\mathbf{g}_0$) and the model ($\mathbf{g}_1$). We will now proceed to approximate that difference with the linear term of the Taylor series:

$$\begin{aligned}\Delta \mathbf{q} &= \mathbf{q}_1 - \mathbf{q}_0 = \mathbf{q}_1 - \mathbf{f}(\mathbf{f}^+(\mathbf{q}_1, \mathbf{g}_0), \mathbf{g}_0) = \mathbf{h}(\mathbf{g}_0) \\ \Delta \mathbf{q} &\approx \frac{\partial \mathbf{h}}{\partial \mathbf{g}}(\mathbf{g}_0 - \mathbf{g}_1) = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{p}} \frac{\partial \mathbf{f}^+}{\partial \mathbf{g}} + \frac{\partial \mathbf{f}}{\partial \mathbf{g}} \right) (\mathbf{g}_1 - \mathbf{g}_0)\end{aligned}\quad (12)$$

Both the partial derivatives $\partial \mathbf{f} / \partial \mathbf{p}$ and $\partial \mathbf{f} / \partial \mathbf{g}$ appearing in (12) are available; it is only $\partial \mathbf{f}^+ / \partial \mathbf{g}$ which requires a separate derivation. Differentiating the first line of (3) over $\mathbf{g}$ gives:

$$\frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} \frac{\partial \mathbf{f}}{\partial \mathbf{g}} + \frac{\partial \mathbf{f}^+}{\partial \mathbf{g}} = \mathbf{0} \Rightarrow \frac{\partial \mathbf{f}^+}{\partial \mathbf{g}} = - \frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} \frac{\partial \mathbf{f}}{\partial \mathbf{g}} \quad (13)$$

Substituting (13) to (12) yields the sought expression for $\Delta \mathbf{q}$:

$$\Delta \mathbf{q} = \left(- \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} \frac{\partial \mathbf{f}}{\partial \mathbf{g}} + \frac{\partial \mathbf{f}}{\partial \mathbf{g}} \right) (\mathbf{g}_1 - \mathbf{g}_0) = (\mathbf{I} - \mathbf{A}\mathbf{G}) \frac{\partial \mathbf{f}}{\partial \mathbf{g}} (\mathbf{g}_1 - \mathbf{g}_0) \quad (14)$$

We will now introduce the following symbols for frequently occurring expressions:

$$\mathbf{L} = \frac{\partial \mathbf{f}}{\partial \mathbf{g}}, \quad \mathbf{H} = \mathbf{I} - \mathbf{A}\mathbf{G}, \quad \Delta \mathbf{g} = \mathbf{g}_1 - \mathbf{g}_0 \quad (15)$$

Please note that $\mathbf{L}$ can be thought of as a **sensitivity matrix**. With the new nomenclature, our calibration equation (14) acquires its final form:

$$\Delta \mathbf{q} = \mathbf{H} \mathbf{L} \Delta \mathbf{g} \quad (16)$$

As the manipulator performs a sequence of movements, all the measurement values $\mathbf{q}_1$ are registered and anticipated sensor readings $\mathbf{q}_0$ are computed for each point, according to (11). Consequently, each point contributes equations of the type (16). Each of those equations relates the geometry errors to the observed discrepancies between true and anticipated sensor readings. Calibration is accomplished though a least-squares fit to the system of all the available equations.

Since, in general, both $\mathbf{H}$ and $\mathbf{L}$ are functions of $\mathbf{g}$, the entire calibration fit has to be iterated a number of times until the fit residue decreases below a desired level. Furthermore, both $\mathbf{H}$ and $\mathbf{L}$ are functions of $\mathbf{p}$, for which, in every iteration, we substitute the result of our increasingly more accurate forward transformation.

5. Redundancy Matrix

It is interesting to note that the key matrix in our calibration formalism (16), $\mathbf{H}$, which we also call **redundancy matrix**, has a rank equal to the number of redundant sensors in the system, since it is a null space of both $\mathbf{A}$ and $\mathbf{G}$:

$$\begin{aligned}\mathbf{H}\mathbf{A} &= (\mathbf{I} - \mathbf{A}\mathbf{G})\mathbf{A} = \mathbf{A} - \mathbf{A}(\mathbf{G}\mathbf{A}) = \mathbf{A} - \mathbf{A}\mathbf{I} = \mathbf{0} \\ \mathbf{G}\mathbf{H} &= \mathbf{G}(\mathbf{I} - \mathbf{A}\mathbf{G}) = \mathbf{G} - (\mathbf{G}\mathbf{A})\mathbf{G} = \mathbf{G} - \mathbf{I}\mathbf{G} = \mathbf{0}\end{aligned}\quad (17)$$

Another interesting observation arises from differentiating the second line of (3) over $\mathbf{q}$:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{p}} \frac{\partial \mathbf{f}^+}{\partial \mathbf{q}} = \mathbf{A}\mathbf{G} \neq \mathbf{I} \Rightarrow \mathbf{H} = \mathbf{I} - \mathbf{A}\mathbf{G} \neq \mathbf{0} \quad (18)$$

In general, $\mathbf{H}$ is a non-zero matrix; it degenerates to a zero matrix only for a non-redundant system.

6. Simulation Example

We have demonstrated the operation of our method in a Mathematica simulation on a simple example involving a 6 DoF motion stage, whose position is measured by 8 displacement sensors interrogating 3 metrology surfaces on the stage (Figure 1). The stage has the following ranges of motion:

$$\begin{aligned}\mathbf{X}, \mathbf{Y}, \mathbf{Z}: & \quad \pm 125 \text{ mm} \\ \mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z: & \quad \pm 0.1 \text{ rad}\end{aligned}$$

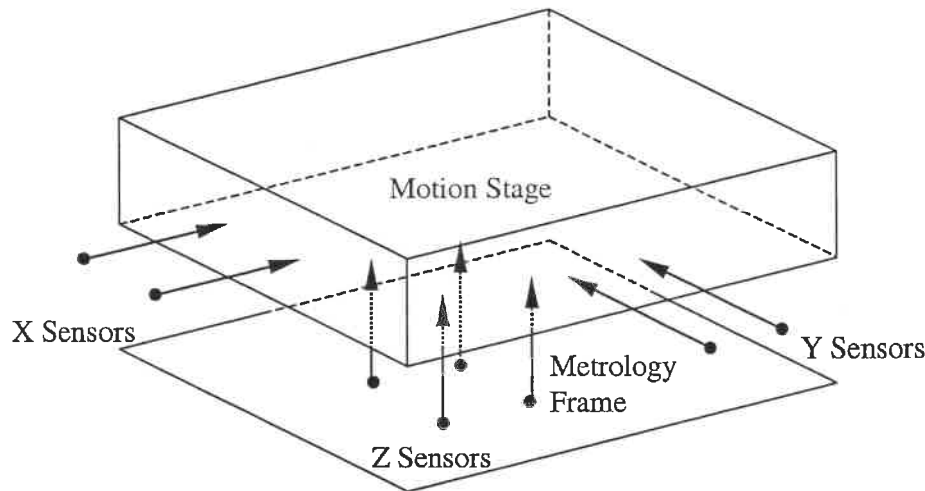


Figure 1 Schematic layout of the idealised motion stage chosen for simulating our calibration method. All the metrology surfaces of the motion stage, as well as all the sensor locations are encumbered with errors, which we are able to calibrate with our method.

The sensors are idealised as providing absolute distances between certain points stationary in the reference frame and the metrology surfaces, which are idealised as planes associated with the moving stage. There are 2 sensors approximating the X coordinate of the stage, 2 for the Y coordinate and 4 for the Z coordinate. This arrangement gives us $8 - 6 = 2$ redundant measurements. The inclination of metrology surfaces and sensor locations in 3 dimensions are initially unknown. Thus the system includes a total of $8 \times 3 + 3 = 27$ geometry parameters. The tolerance which we assume for simulating the translational errors amounts to 1 mm, the tolerance for angular errors is 0.01 rad. The starting iteration assumes a nominal layout. In the course of the simulation, the stage is brought to $6^3 = 729$ different positions representing all possible combinations of the extreme negative, zero and the extreme positive value of each of the 6 DoF.

The simulation shows that all the geometry parameters can be calibrated in-situ, utilising only the internal displacement sensors. The entire process requires 5 iterations, while the condition numbers of the calibration matrix are on the order of 10^6 . The calibration matrix has 12 zero singular values, which correspond to the arbitrary locations of coordinate frames used for defining the stage geometry. Those frame locations can be constrained artificially with additional equations or left to the matrix pseudo-inversion, which will constrain them naturally.

When controlled with the parameters obtained from calibration, the manipulator is capable of accurate movements up to the constant spatial offset at both coordinate frames used for defining positions.

7. Summary

Our paper presents a methodology of calibrating the geometry of a redundant position measurement system in situ. We introduce the notions of inverse and forward coordinate transformation associated with the degrees of freedom we seek to control and the sensor readings at our disposal. We subsequently propose to apply a composition of both the transformations to sensor readings obtained at various manipulator poses and prove that the result of that composition is algebraically related to the mismatch between the manipulator and its model. Crucially, we introduce the notion of the redundancy matrix $\mathbf{H}$ governing the manner in which the mismatch between reality and our model becomes visible to our sensors. In conclusion, as a proof of principle, we simulate an idealised 6 DoF motion stage equipped with 8 displacement sensors and succeed in calibrating its complete geometry up to constant spatial offsets. We believe that the methodology we present is applicable to a wide range of manipulators and can be extended theoretically in many directions.

A SYSTEMATIC APPROACH TO THE MODELLING OF MEASUREMENTS FOR UNCERTAINTY ANALYSIS

Klaus-Dieter Sommer¹, Bernd R.L. Siebert², Manfred Kochsiek² and Albert Weckenmann³

¹Landesamt fuer Mess- und Eichwesen Thueringen (LMET); Unterpörlitzer Strasse 2, D-98693 Ilmenau, Germany

²Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, D-38116 Braunschweig, Germany

³Friedrich-Alexander-University Erlangen-Nuremberg, Naegelsbachstrasse 25, D-91052 Erlangen, Germany

1 INTRODUCTION

The modern evaluation of measurement uncertainty is based on both, the knowledge about the measuring process and the (input) quantities that may influence the measurement result. The knowledge about these input quantities is to be expressed by means of appropriate probability distributions functions (pdf) for these quantities, whereas the knowledge about the measuring process is to be condensed to the so-called model equation. This equation mathematically interrelates the measurement Y and all relevant input quantities $X_1, \dots, X_N$:

$$Y = f_Y(X_1, \dots, X_N). \quad (1.1)$$

Since neither the *Guide to the Expression of Uncertainty in Measurement (GUM)* [1] nor other relevant uncertainty documents provide sufficient guidance on systematic modelling procedures to practitioners, modelling appears to be the most difficult problem in uncertainty evaluation. First approaches to a systematic and *GUM*-consisting modelling procedure were made by *Bachmair* [2], *Kessel* [3], *Kind* [4] and *Sommer et al.* [5; 6]. This paper reports the progress made and extends the previous work.

2 BASIC RELATIONSHIPS

The *GUM* concept for evaluating the uncertainty is based on the knowledge about the measuring process and the quantities which may have influence on the measurement result and their associated uncertainties. In accordance with the *GUM* concept [1] the (unavoidably incomplete) knowledge about each contributing input quantity is to be expressed by means of pdfs $g_{X_i}(\xi_i)$. The expectation value of the pdf is the best estimate of the value of the quantity,

$$x_i = E[X_i] = \int_{-\infty}^{+\infty} g_{X_i}(\xi_i) \xi_i d\xi_i, \quad (2.1)$$

and its standard deviation is the uncertainty u_{X_i} associated with this estimate,

$$u_{X_i} = \left\{ E[(X_i - x_i)^2] \right\}^{1/2} = \left\{ \int_{-\infty}^{+\infty} g_{X_i}(\xi_i) (\xi_i - x_i)^2 d\xi_i \right\}^{1/2} \quad (2.2)$$

Lower case Greek letters are used for the possible values of a quantity, ξ for X and η for Y .

The pdfs for the input quantities are obtained from given information by utilizing the principle of maximum information entropy (pme) [7]. New information, e. g. from additionally measured data is included by using the *Bayes* theorem [8]. The standard-*GUM* procedure condenses the pme and parts of the *Bayesian* concept to the evaluation methods of *type-A* (statistical analysis of information from repeated measurements) and *type-B* (evaluation by any other means). In both cases, one obtains unambiguously a pdf for each input quantity.

The pdf for the measurand Y , $g_Y(\eta)$, is given by the integral

$$g_Y(\eta) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g_{X_1, \dots, X_N}(\xi_1, \dots, \xi_N) \delta(\eta - f_Y(\xi_1, \dots, \xi_N)) d\xi_1 \dots d\xi_N \quad (2.3)$$

where f_Y is the functional relationship of the values of the involved quantities, ξ , with the respective value of the measurand, η . From the above pdf $g_Y(\eta)$, the expectation value of the measurand $y = E[Y]$ and its associated uncertainty u_y can be calculated as follows:

$$y = \int_{-\infty}^{+\infty} g_Y(\eta) \eta d\eta, \text{ and } u_y = \left\{ \int_{-\infty}^{+\infty} g_Y(\eta) (\eta - y)^2 d\eta \right\}^{1/2} \quad (2.4) \quad (2.5)$$

This way of uncertainty determination by means of pdf propagation is illustrated in Fig. 1. But because equation (2.3) can analytically be computed in fairly simple cases only, modern uncertainty evaluation uses the *Monte-Carlo Method* as an integration technique [9]. *Monte-Carlo* techniques are based on the sampling of the cumulative input pdfs: With uniform probability, probability values $G_{X_i}(\xi_i)$ of the input quantities are selected. Then, the related arguments ξ_i of each quantity are to be combined in accordance with the model equation for the values, $y = f_Y(\xi_1, \dots, \xi_N)$, where f_Y is identical with f_Y in equation (1.1) yielding values η_k . From a sufficient number of samples, a frequency distribution is obtained that approximates the probability distribution $g_Y(\eta)$ Fig. 2 illustrates the *Monte-Carlo* integration technique that handles linear and non-linear models alike.

In contrast to *Monte-Carlo* techniques, the standard-*GUM* procedure (see Fig. 3) [1] applies to linear or linearized models only. In practice, most dependencies on input quantities can be approximated by first-order *Taylor* series expansion within the range extended by the uncertainties associated with the values of the input quantities:

$$\eta = f(\xi) \cong f(x) + \sum_{i=1}^N c_i (\xi_i - x_i) \quad (2.6)$$

where η denotes a possible value of the output quantity Y , and ξ is used as abbreviation for $\xi_1, \dots, \xi_N$, and x for $x_1, \dots, x_N$. The c_i are called sensitivity coefficients and given by

$$c_i = \frac{\partial f(X_1, \dots, X_N)}{\partial X_i} \Big|_{\forall \xi_i = x_i}. \quad (2.7)$$

Therefore, for a linear (or linearized) model equations, the expectation value of the output quantity Y is

$$Y = E[X_i] = f_Y(x_1, \dots, x_N). \quad (2.8)$$

The uncertainty associated with this expectation is obtained from the law of *Gaussian* uncertainty propagation:

$$u_y = \left\{ \sum_{i=1}^N c_i^2 u_{x_i}^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N c_i c_j u_{x_{ij}} \right\}^{1/2}, \quad (2.9)$$

where $u_{x_{ij}} = u_{x_i} u_{x_j} \cdot r_{ij}$ is the covariance of the quantities X_i and X_j , and r_{ij} the correlation coefficient.

From the above basic relationships on uncertainty evaluation, it may be concluded that – independent on the method utilized – the model equation is an indispensable prerequisite of modern uncertainty evaluation.

3 MODELLING PROCEDURE

3.1 Concept

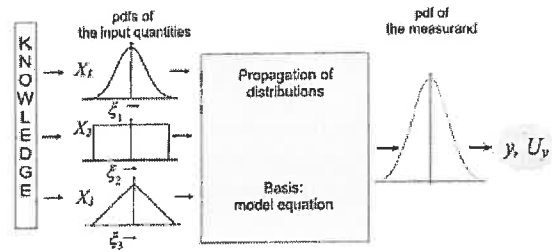


Fig. 1: Illustration of the concept of pdf propagation in uncertainty evaluation [5]. Symbols: Y – Measurand; $X_1, \dots, X_N$ – Input quantities; $y = E[Y]$ – expectation value of the measurand; u_y – uncertainty associated with y

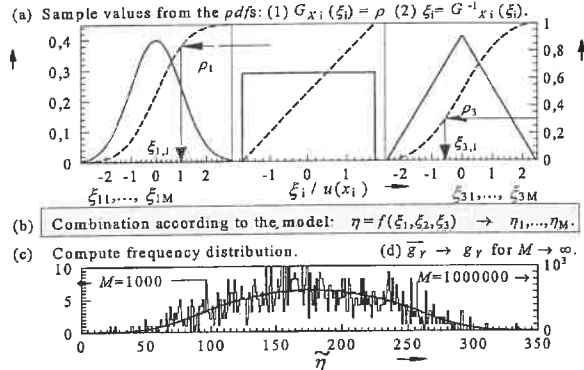


Fig. 2: Illustration of Monte-Carlo integration (a) The upper three graphs show (solid lines) the pdfs to the (input) quantities and (dashed lines) the respective cumulative distributions. (b) Combination of the arguments (c) The curves in the bottom part show computed frequency distributions for two sample sizes M . (d) Approximation of the pdf to the measurand

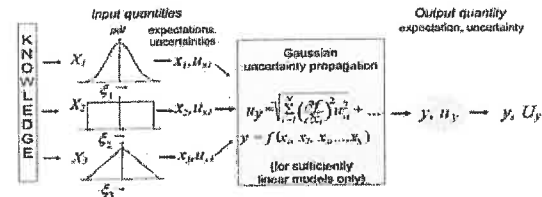


Fig. 3 Illustration of the concept of the standard-*GUM* procedure [1; 5]. Symbols see Fig. 1

The modelling concept presented [5; 6] is based on the idea of the *measuring chain* which constitutes the path of the measurement signal from cause to effect. In metrological practice, the following assumptions can be made:

- At least in narrow ranges around the operating points, the great majority of measuring systems and devices (elements) may be regarded to have linear characteristics and can be approximated by first-order Taylor series expansion respectively (see equation (2.6.)).
- The (steady-state) transmission behaviour of a measuring system is always related to well-adjusted and known operating conditions.
- The „real world of measurement“ may be taken into consideration by means of deviations of the real influence quantities and other parameters.

On the above assumptions, in steady state, almost all functional elements or operational steps of a measuring system or process may be described by an approximately constant transmission factor and by *deviations* representing the imperfection of the measurement. Deviations may have impact on transmission factors and they may result in offsets of the outputs. Fig. 4 illustrates this concept of a perturbed transmission element that mathematically can be expressed by the following relationship:

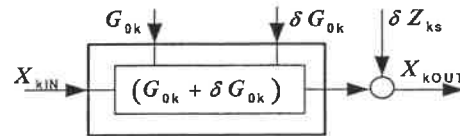


Fig. 4: Concept of a perturbed transmission element [5]

$$X_{KOUT} = X_{KIN} (G_{OK} + \delta G_{OK}) + \delta Z_K \quad (3.1)$$

where: X_{KIN} – quantity acting at the input of the element k ; X_{KOUT} – quantity at the output of the element k ; G_{OK} – transmission factor; δG_{OK} – parameter deviation δZ_K – parameter deviation.

The above concept allows to graphically describe the cause-and-effect relationship of a measurement at any operating point.

3.2 Standard modelling components

For the required graphical depiction of the cause-and-effect relationships of the measurements to be modelled (see 3.3), only three types of standard modelling components are employed:

- **Parameter sources** to provide or reproduce a measurable quantity.
- **Transforming units** that represent any kind of parameter processing and influencing.
- **Indicating units** to indicate their input quantities.

3.3 Modelling procedure

Based on the above presumptions, a straightforward and highly versatile modelling procedure has been developed [5; 6] that consists of the following elementary steps:

- 1st step:** Description of the measurement, analysis of the measuring process by decomposing it in terms of standard modelling components (see 3.2) with a view to form the measuring chain.
- 2nd step:** Graphical depiction of the cause-and-effect relationship for a fictitious ideal (unperturbed) measurement in terms of standard modelling components.
- 3rd step:** Inclusion of all imperfections and effects that may perturb the fictitious ideal measurement, such as, for example, influences, incomplete knowledge about parameters etc., and representation of the resulting cause-and-effect relationship graphically and, in turn, mathematically for the real measurement.
- 4th step:** Identification and inclusion of possible correlation in the measuring chain [10]:
1st way: If correlation is caused by conjoint functional dependencies on a third quantity, such as, for example, on temperature, these dependencies are to be accounted. The way to do this is to introduce these

dependencies in the graphical and mathematical cause-and-effect relationship and, therewith, to dissolve correlation. If possible to go, this first way is to be preferred.

2nd way: Correlation is taken into account in accordance with the law of *Gaussian* uncertainty, propagation (see equation (2.9)). This way, however, requires the knowledge of the (estimated or experimentally determined) value of the correlation coefficient [10].

5th step: Inversion of the cause-and-effect relationship to derive explicitly the mathematical relationship between the output quantity and the relevant input quantities.

4 EXAMPLE

4.1 Modelling procedure

The modelling procedure is explained with the (simplified) example of the determination of a weighing value by direct weighing.

1st step:

- *Description of the measurement:* The weighing value of a weight of about 10 kg is to be determined by direct weighing taking one reading. The scale used is a 12 kg- range instrument (scale value: 0.1 g; maximum permissible error on verification, mpev = 0.5 g).
- *Measurand:* Weighing value W_X of the weight piece.
- *Measurement method:* Direct measurement.
- *Analysis of the measuring process:* The weight piece may be regarded as the parameter source, and the imperfect „coupling“ of the unknown weighing value with the scale (causes: air buoyancy, magnetic, susceptibility etc.) may be described by a transforming unit. The scale itself may be represented by an indicating unit.

2nd step: Cause-and-effect relationship of the fictitious ideal measurement:

Fig. 6 shows the graphical cause-and-effect relationship of the idealized (unperturbed) measurement.

3rd step: Cause-and-effect relationship of the real measurement:

Fig. 7 shows the graphical cause-and-effect relationship. The following imperfections have been introduced:

- δW_{CPL} – deviation due to eccentric loading, vibration, air convection, magnetic susceptibility etc.;
- k_B – $k_B = (1 - \rho_a \rho_X^{-1})(1 - \rho_{1,2} \rho_{8000}^{-1})^{-1}$, air buoyancy factor, where ρ_a is the air density ρ_X is the density of the weight to be calibrated, and $\rho_{1,2} = 1.2 \text{ kgm}^{-3}$, $\rho_{8000} = 8000 \text{ kgm}^{-3}$;

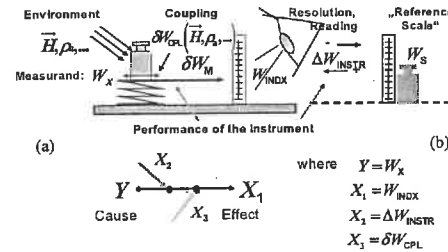


Fig. 5: (a) Simplified example: Direct determination of an unknown weight; (b) Visualisation of the cause-and-effect relationship by means of a graph. Symbols: W_X - measurand; δW_{CPL} - unknown deviation due to the imperfect „coupling“ of the measurand with the scale (causes: air buoyancy etc.); W_{IND} - indicated quantity; ΔW_{INSTR} - instrumental error of the weighing instrument; W_S - standard weight; $\vec{H}$ - magnetic field strength; ρ_a - air density

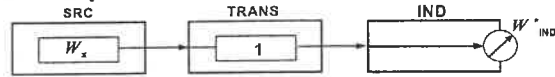


Fig. 6: Graphical depiction of the cause-and-effect relationship of the fictitious ideal weighing procedure in accordance with 4.1. Symbols: W_X - measurand; W_{IND}^* - indicated quantity

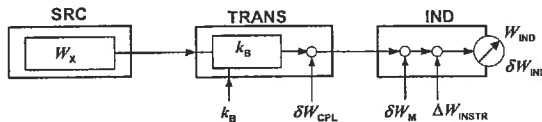


Fig. 7: Cause-and-effect relationship of the real measurement in accordance with the described example (clause 4.1) Symbols: see text

- $\delta W_M(t_a)$ – deviation due to the susceptibility of the weighing instrument to ambient temperature t_a ;
- ΔW_{INSTR} – instrumental error of the weighing instrument;
- δW_{IND} – deviation due to the limited resolution of the instrument.

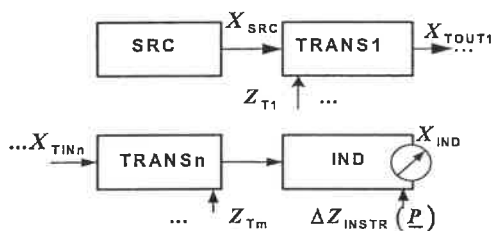


Fig. 8: Generic structure of the cause-and-effect relationship of direct measurements. Symbols: X_{SRC} – output quantity of the parameter source; $\underline{P}$ – measurement conditions; $Z_{T1}, \dots, Z_{Tm}$ – additional input quantities, such as, for example transmission factors and deviations; X_{IND} – indicated quantity, $\Delta Z_{INSTR}(\underline{P})$ instrumental error;

In mathematical terms, the cause-and-effect relationship of the real measurement reads:

$$W_{IND} = W_X k_B + \delta W_{CPL} + \delta W_M + \Delta W_{INSTR} + \delta W_{IND} \quad (4.1)$$

4th step: Correlation:

For the sake of simplification, all involved quantities, parameters and observations are assumed to be independent of each other.

5th step: Model equation:

From the cause-and-effect relationship of the real measurement (see equation (4.1)), the following model equation is obtained:

$$W_X = (W_{IND} - \delta W_{IND} - \Delta W_{INSTR} - \delta W_M - \delta W_{CPL}) k_B^{-1} \quad (4.2)$$

4.2 Evaluating the measurement uncertainty

Due to the almost linear model equation of the chosen example (see equation (4.2)), the standard-*GUM* Method may be used to determine the measurement uncertainty. After modelling the measurement, the most important step is to evaluate the involved input quantities W_{IND} , δW_{IND} , ΔW_{INSTR} , δW_M , δW_{CPL} , k_B . To each of these quantities, an

appropriate pdf is to be assigned. W_{IND} will be clearly indicated. The values of δW_{IND} may be derived from the instrument's resolution and for ΔW_{INSTR} from the maximum permissible error stated for the instrument. The knowledge about δW_{CPL} and δW_M may be taken up from the manufacturers manual or from the requirements set up in the European standard *EN 45501*. k_B is to be estimated based on the knowledge about the ambient conditions and about the weight piece to be measured.

5 ROLE OF THE MEASUREMENT METHOD

The structure and the chaining sequence of the cause-and-effect relationship are determined by the *method of measurement* used. Direct measurements result in an unbranched chain of the components utilized (see Fig. 8).

Other methods are used to achieve higher accuracies and to ensure proper traceability of calibration results. These methods mostly result in branched cause-and-effect relationships. Examples are given with the direct comparison of two indicating measuring instruments and substitution method. Fig. 9 and 10 show the generic structures of their cause-and-effect relationship. When deriving the mathematical cause-and-effect relationship from block diagrams having branched structures, such as, for example, the above methods, for each branch a separate (partial) equation is to be derived [5]. Thereby, influences and imperfections of the commonly used parts of the paths should be assumed to be correlated.

6 CONCLUSION

Although it seems not possible to develop a theory that allows for a stringent construction of a model, it is, nevertheless, possible to achieve systematic modelling based on the presented concept. Systematic modelling may be

seen as an important improvement of uncertainty evaluation. The particular concept presented is applicable to most areas of uncertainty evaluation of measurements performed. It is clearly structured in five elementary steps, and only three types of standard modelling components are employed.

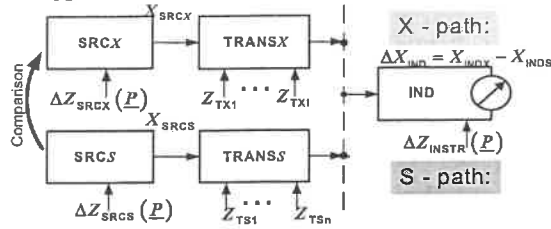


Fig. 9: Generic structure of the cause-and-effect relationship of a calibration of a material measures employing *substitution method*. Symbols: SRCX – material measure to be calibrated; SRCS – standard material measure; X_{SRCX} – output quantity of SRCX; $\Delta Z_{SRCX}(P)$ – measurand; P – calibration conditions; X_{SRCS} – output quantity of the standard; $\Delta Z_{SRCS}(P)$ – instrumental error of the standard used; TRANSX / TRANSS – transforming units; IND – indicating instrument (comparator); $\Delta Z_{INSTR}(P)$ – instrumental error of the comparator; $Z_{TX1}, \dots, Z_{TXI}$ and $Z_{TS1}, \dots, Z_{TSn}$ – additional input quantities; $\Delta X_{IND} = X_{INDX} - X_{INDS}$ – indication difference between the material measure and the standard

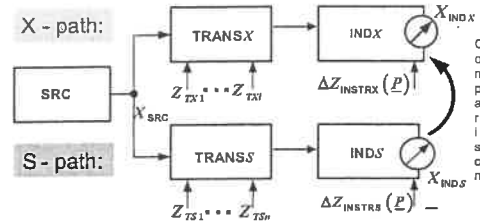


Fig. 10: Generic structure of the cause-and-effect relationship of a calibration employing the *direct comparison method* of two indicating measuring instruments. Symbols: X_{SRC} – quantity provided by a parameter source SRC; TRANSX – transforming unit of the X-path; TRANSS – reference transforming unit; INDX – instrument under test; INDS – indicating standard; $Z_{TX1}, \dots, Z_{TXI}$ and $Z_{TS1}, \dots, Z_{TSn}$ – additional input quantities; $\Delta Z_{INSTRX}(P)$ – instrumental error of the instrument under test (measurand) depending on the calibration conditions P ; $\Delta Z_{INSTRS}(P)$ – instrumental error of the standard used; X_{INDX} – indication of the instrument under test; X_{INDS} – indication of the standard

REFERENCES

- 1 Guide to the Expression of Uncertainty in Measurement (GUM), first edition, 1993, corrected and reprinted 1995, International Organization for Standardization (ISO), Geneva, 1993
- 2 H. Bachmair: A simplified method to calculate the uncertainty of measurement for electrical calibration and test equipment. Proceedings of the 9th INTERNATIONAL METROLOGY CONGRESS, Bordeaux, 18 - 21 October 1999, Mouvement Francais pour la Qualite, Nanterre Cedex 1999 (ISBN 2-909430-88x)
- 3 W. Kessel: Messunsicherheit - einige Begriffe und Folgerungen fuer die messtechnische Praxis. PTB-Mitteilungen, vol. 111, 3, pp. 226-244, 2001, Wirtschaftsverlag NW, Bremerhaven 2001 (ISSN 030-834X)
- 4 D. Kind: Die Kettenschaltung als Modell zur Berechnung von Messunsicherheiten. see [3], pp. 338-341
- 5 Sommer, K.-D.; Kochsiek, M.; Siebert B.R.L.; Weckenmann, A.: A Generalized Procedure for Modelling of Measurement for Evaluating the Measurement Uncertainty. 17th IMEKO WORLD CONGRESS 2003, Dubrovnik, Proceedings, IMEKO and Croatian Metrology Society, Zagreb 2003 (ISBN 953-7124-00-2)
- 6 Sommer, K.-D.; Kochsiek, M.; Weckenmann, A.: Modelling of Temperature and Pressure Sensor Calibration for Evaluating the Uncertainty. MSC Measurement Science Conference 2004, Anaheim, California, Conference Proceedings, Measurement Science Conference, Newport Beach, California 2004
- 7 W. Weise, W. Woeger: Messunsicherheit und Messdatenauswertung, WILEY-VCH, Weinheim, 1999
- 8 Silvia, D.S.: Data Analysis - A Bayesian Tutorial. Clarendon Press, Oxford 1996
- 9 Siebert, B.R.L.; Sommer, K.-D.: Weiterentwicklung des GUM und Monte-Carlo-Techniken. tm Technisches Messen 71 (2002), Heft 2 (February 2004), S. 67 - 80 (ISSN 0178-2312)
- 10 Sommer, K.-D.; Kochsiek, M.; Siebert, B.R.L.: A Consideration of Correlation in Modelling and Uncertainty Evaluation of Measurement. NCSL International Workshop and Symposium, Salt Lake City, Conference Proceedings, NCSL International, Boulder, Colorado 2004
- 11 Sommer, K.-D.; Kochsiek, M.; Siebert, B.R.L.: Modelling of Weighing Procedures for Uncertainty Evaluation. South Yorkshire International Conference 2003, Barnsley. Conference Proceedings, South Yorkshire Trading Standards Unit, Chapelton, Sheffield 2003

Application of Simulation Software to Coordinate Measurement Uncertainty Evaluation

Kim D. Summerhays, Jon M. Baldwin, Daniel A. Campbell and Richard P. Henke,
MetroSage LLC, 26896 Shake Ridge Road, Volcano, CA 95689

Introduction Measurement uncertainty plays a vital role in establishing traceability to national and international standards [1]. Likewise, it is a critical factor in the arbitration of issues pertaining to product conformance [2]. For dimensional metrology, it is necessary to associate with each GD&T parameter an assessment of its uncertainty at a specified level of confidence. As a result, for manufacturers to adhere to current standards, inspection reports must be supplemented with statements like, "The uncertainty of the diameter of this nominally 6 mm diameter hole is 0.02 mm, at 95% confidence." This is an example of what is often called a *task-specific* measurement uncertainty statement.

The available methods for CMM uncertainty evaluation have been summarized in a draft ISO technical report [3]. They are:

1) Sensitivity Analysis, which involves listing each uncertainty source, its magnitude, effect on the measurement result and its correlation with other uncertainty sources, then combining them in a manner that accounts for the sensitivity factor of each source and the interactions between sources. This is the approach described in the ISO Guide to Uncertainty in Measurements (GUM) and is particularly useful if a mathematical model of the measuring process can be had, because direct computation of the sensitivity coefficients is then possible.

2) Expert Judgment, which may be the only available method if a mathematical model or measurement data are not available. Its limitations in producing a defensible uncertainty statement are evident.

3) Substitution, wherein repeated measurement of a calibrated master part yields a range of errors and thus the uncertainty. This is a powerful method of capturing the relevant error sources and their interactions. Its major disadvantages are expense (need for multiple master parts) and a reduction of the range of utility of the CMM.

4) Computer Simulation, where a virtual model of the CMM system is created, including the error sources. Examining the task-specific results yields estimates of both bias and measurement variability and hence uncertainty.

5) Measurement History, which is useful if large numbers of measurements over time are available. This method can place an upper bound on measurement uncertainty. It fails to detect measurement bias.

In considering these alternative methods, the ISO draft [3] points out that,

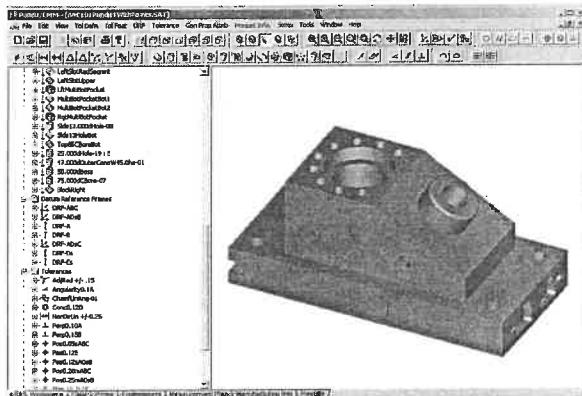
The benefit of computer simulation is derived from repeated simulations of different measurement scenarios, where each scenario involves a specific set of measurement errors (as opposed to uncertainties). The use of specific measurement errors, together with the mathematical model, often allows a more complete description of the interactions, i.e. correlations, between sources than attempting to calculate sensitivity coefficients. (In some cases sensitivity coefficients are impossible to calculate analytically since the measurement process cannot be analytically described).

For credibility, whatever method chosen must be comprehensive i.e. all the major influence variables must be considered and all necessary GD&T parameters must be supported. It must produce accurate and reliable results, and the evaluations should conform to established GD&T parameter definitions. Several other qualities are highly desirable: It should be versatile by supporting a useful variety of CMM and probing system error models, and workpiece and CMM thermal models. It should demonstrate fidelity by allowing realistic construction of measurement scenarios, metrology hardware configurations, and correct choice of geometric fitting algorithms. It should be interoperable by accepting data from legacy sources, e.g. existing workpiece designs and inspection programs, and should provide a defined interface for communicating uncertainty information with other applications. It should be flexible, offering a spectrum of tradeoffs between cost of system characterization and quality of the resulting uncertainty evaluations.

PUNDIT/CMM™ We have developed a convenient tool, PUNDIT/CMM [4], for the evaluation of CMM-based task-specific measurement uncertainties. PUNDIT/CMM addresses all the principal error sources in CMM measurements (CMM, probe, thermal conditions, feature surface characteristics, sampling patterns, etc.) and does so with a modular architecture that facilitates enhancements of the error models as the state of the art advances. It incorporates a full solid model of the workpiece, supporting datum reference frames, tolerancing, and deviations from perfect surface form. As a key principle of operation, it uses the NIST-developed Simulation by Constraints (SBC) statistical methodology [5,6]. SBC allows for CMM and probe modeling based on standard (e.g. B89 or ISO 10360) performance test results by establishing ensembles of virtual CMM and probe states compatible with specific test result values. Then, using specified thermal conditions, sampling strategies and analysis algorithms, we repeatedly simulate workpiece inspection, based on a randomized treatment of these states. This yields, for each tolerated feature characteristic, a distribution of measurement results, showing measurement bias and variability. The activities associated with these evaluations are divided into seven logical categories: workpiece definition, manufacturing parameters, CMM definition, sensor (probe) definition, environmental effects, the measurement plan, and analysis and results. These, with the exceptions of the first and last categories, can be dealt with in any convenient order. Each is accessed by selecting the corresponding “tab” in the PUNDIT/CMM user interface, as discussed below.

A Quick Tour of PUNDIT/CMM™

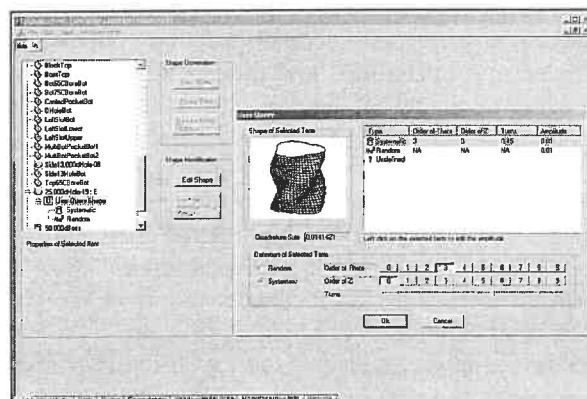
Workpiece and Tolerancing Definition



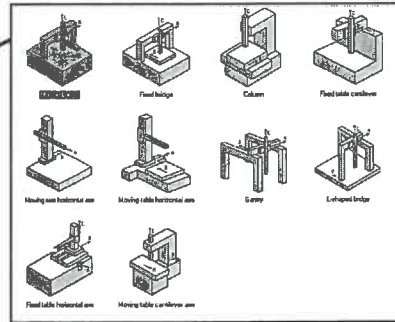
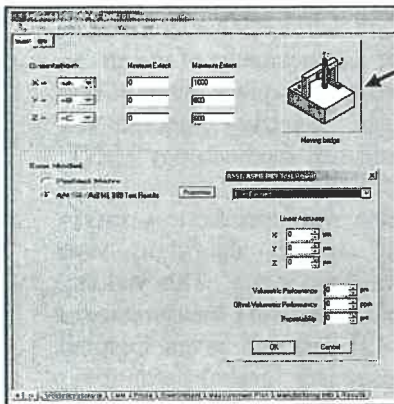
In the Workpiece Tab, the user creates or imports a solid model representation of the workpiece. (Import options: ACIS, CATIA, Pro/E, STEP, and SolidWorks.) PUNDIT/CMM recognizes tolerance features and allows definition of datums and datum reference frames and the establishment of specific tolerances for feature location, orientation, size and form. Moreover, it can *check* the tolerancing to verify that it is complete, consistent and unambiguous. Problems encountered here are clearly identified for design phase correction, saving time and money later in the product life-cycle.

Manufacturing Information for Feature Form Errors

Under the Manufacturing Info Tab, the user can, if he/she desires, assign specific types of systematic and/or random form errors to feature surfaces. These may be of classical analytic form or derived from dense measurement of representative features produced by particular methods of manufacture. Estimated amplitudes of these errors can also be specified. (The example here is a 3-lobe helical form often found in holes produced by drilling operations.) Systematic feature form errors are known to contribute largely to measurement uncertainty for certain probe sampling patterns. PUNDIT/CMM can predict these contributions.



CMM Definition: Geometry, Working Volume and Error Model

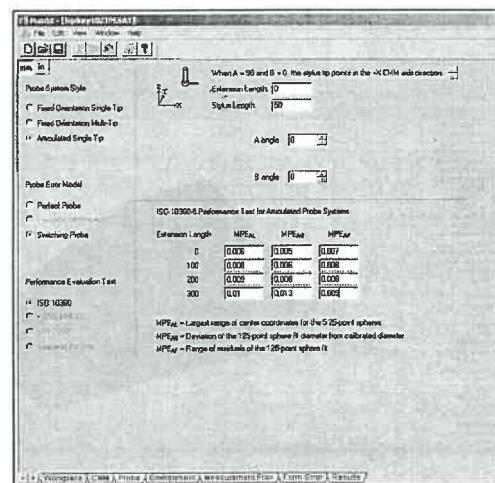


In the CMM Tab, the user identifies the CMM geometry type, the extent of the axes, and the desired error model. Perfect CMM is one option. Data from various alternative suites of CMM performance evaluation tests can also be employed. Such data may be from tests on the user's own CMM or from the CMM manufacturer's data

sheets. Each simulation of workpiece measurement, employs a different "virtual CMM" (i.e. a different set of rigid body errors) that is consistent with the specified test data. These errors then propagate into individual measurement points in the inspection simulations and from there into the toleranced characteristics of the substitute geometry features.

Sensor Definition: Configuration, Error Model and Evaluation

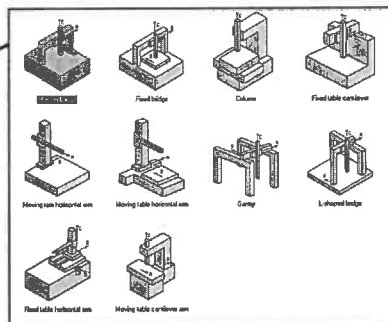
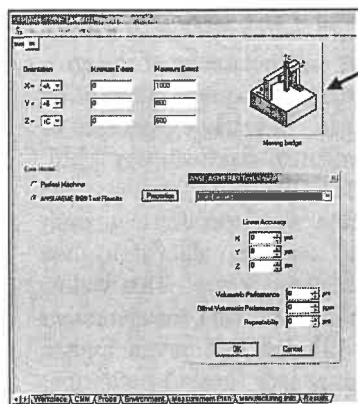
Sensors are defined under the PUNDIT/CMM Probe Tab. The user can select a single-tip fixed probe, a multi-tip fixed probe, or a single-tip articulated probe. Error models vary from perfect probe, to piezoelectric models, to switching probes. A variety of options are available for specification of performance evaluation. These include those mandated in ISO 10360, B89.4.1, and in VDI/VDE 2617. In a way analogous to the treatment of CMMs, the measurement simulations are conducted with the introduction of probe errors on individual sample points. As with CMM errors, these too are propagated through fitting algorithms to yield errors in the final GD&T parameters of interest.



Environmental Effects

Under the Environment Tab, PUNDIT/CMM currently provides a static thermal model for the CMM and both static and dynamic thermal modeling options for the workpiece. The user can stipulate expansion coefficients, temperatures, and their corresponding uncertainties (C_T , T , ΔC_T , and ΔT) for CMM and workpiece in the static mode. The dynamic model allows for the effects of a workpiece whose temperature is varying during measurement, an undesirable circumstance, yet often encountered in the shop-floor environment. Here, the user specifies the order of feature measurement, estimates initial and final workpiece temperatures, and evaluates certain parameters related to rate of measurement. In its simulations, PUNDIT/CMM can emulate various CMM software behaviors for thermal compensations. These include no temperature compensation, CMM compensation only, CMM and workpiece with the workpiece taken as at the same temperature as the CMM, and "full temperature compensation." In this regard, it should be noted that even with "full temperature compensation" in the CMM software, uncertainties in the values of the expansion coefficients (ΔC_T) and temperatures (ΔT) will give rise to contributions to overall measurement uncertainty.

CMM Definition: Geometry, Working Volume and Error Model

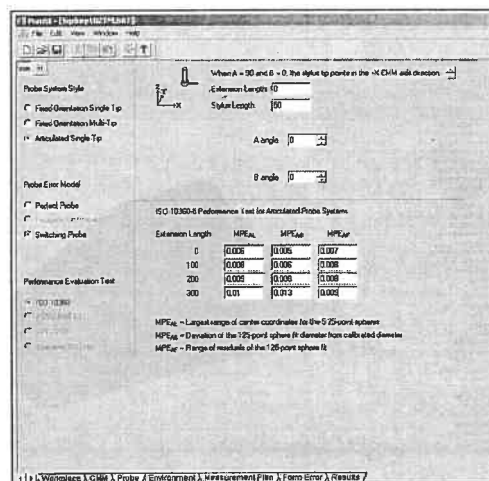


In the CMM Tab, the user identifies the CMM geometry type, the extent of the axes, and the desired error model. Perfect CMM is one option. Data from various alternative suites of CMM performance evaluation tests can also be employed. Such data may be from tests on the user's own CMM or from the CMM manufacturer's data

sheets. Each simulation of workpiece measurement, employs a different "virtual CMM" (i.e. a different set of rigid body errors) that is consistent with the specified test data. These errors then propagate into individual measurement points in the inspection simulations and from there into the toleranced characteristics of the substitute geometry features.

Sensor Definition: Configuration, Error Model and Evaluation

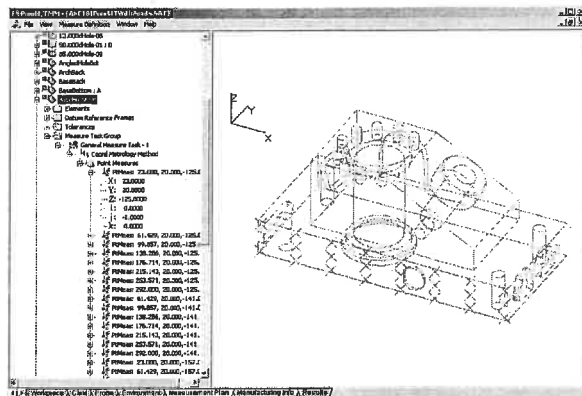
Sensors are defined under the PUNDIT/CMM Probe Tab. The user can select a single-tip fixed probe, a multi-tip fixed probe, or a single-tip articulated probe. Error models vary from perfect probe, to piezoelectric models, to switching probes. A variety of options are available for specification of performance evaluation. These include those mandated in ISO 10360, B89.4.1, and in VDI/VDE 2617. In a way analogous to the treatment of CMMs, the measurement simulations are conducted with the introduction of probe errors on individual sample points. As with CMM errors, these too are propagated through fitting algorithms to yield errors in the final GD&T parameters of interest.



Environmental Effects

Under the Environment Tab, PUNDIT/CMM currently provides a static thermal model for the CMM and both static and dynamic thermal modeling options for the workpiece. The user can stipulate expansion coefficients, temperatures, and their corresponding uncertainties (C_T , T , ΔC_T , and ΔT) for CMM and workpiece in the static mode. The dynamic model allows for the effects of a workpiece whose temperature is varying during measurement, an undesirable circumstance, yet often encountered in the shop-floor environment. Here, the user specifies the order of feature measurement, estimates initial and final workpiece temperatures, and evaluates certain parameters related to rate of measurement. In its simulations, PUNDIT/CMM can emulate various CMM software behaviors for thermal compensations. These include no temperature compensation, CMM compensation only, CMM and workpiece with the workpiece taken as at the same temperature as the CMM, and "full temperature compensation." In this regard, it should be noted that even with "full temperature compensation" in the CMM software, *uncertainties* in the values of the expansion coefficients (ΔC_T) and temperatures (ΔT) will give rise to contributions to overall measurement uncertainty.

Measurement Plan: Sampling Patterns, Sensor Usage, Workpiece Position in CMM

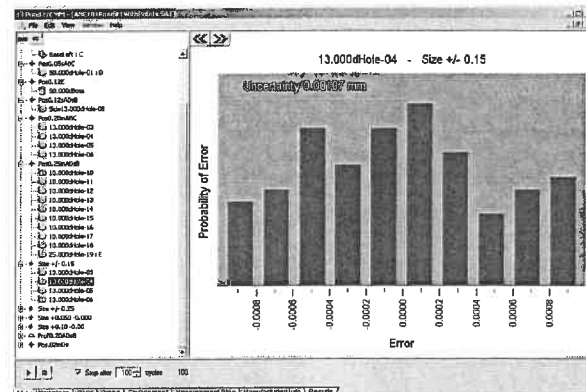


In the Measurement Plan Tab, the user stipulates the number and location of sampling points for features, the sensor(s) to be used in gathering that data, the location and orientation of the workpiece in the CMM working volume and, if needed, the order of feature measurement. Convenient “wizards” are available to automate the sample point specification process. Regular rectangular (eclipsed) or staggered patterns of rows and columns can be selected, or point density methods may be used. Custom patterns may be created. The user can specify offsets from edges and any points falling into voids on feature surfaces are

automatically discarded. Patterns can be easily edited and convenient three-dimensional displays show clearly where sample points are located.

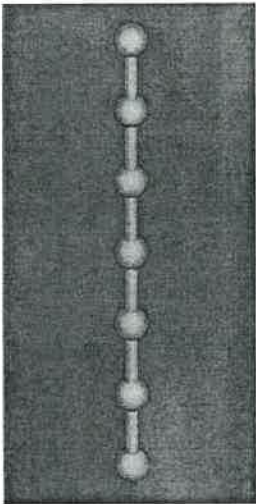
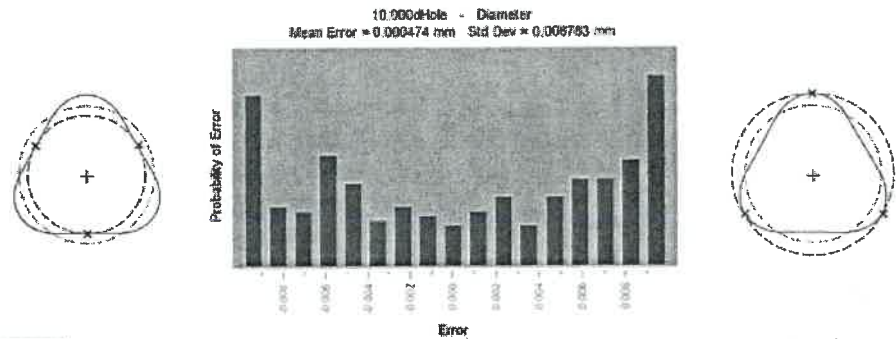
Analysis and Results: Task-Specific Measurement Bias, Variability & Uncertainty Evaluation

When all the influence parameters have been specified, the user can then move to the Results Tab, select the number of simulation cycles to run, and launch the simulation process. Every toleranced feature characteristic yields a histogram of results from which the measurement uncertainty can be inferred. Convergence generally occurs in one or a few hundred cycles, taking a matter of seconds to a few minutes on a Pentium class processor, depending on the complexity of the workpiece. A summary report, in a standard format easily read by other applications, is exportable on demand.



Example Applications: “What if?” Scenarios PUNDIT/CMM provides defensible, real-world measurement uncertainty evaluations. But there is more. One of the many advantages of measurement simulators is the ability to carry out “What if?” calculations. These serve to aid in the verification of the software itself, by allowing isolation of individual effects for study. They also provide for assessment of which aspects of the measurement operation may be dominant contributors to uncertainty in the results. Moreover, “What if?” experiments can provide insights useful in the training of users in proper CMM measuring procedures. As examples, consider first a 10 mm ID cylinder sampled at 3 equiangular points at each of 3 levels along its length (the points at each level eclipse each other when the cylinder is viewed end-on). The cylinder surface is specified to have a 3-lobe (sinusoidal) form error (i.e. cylindricity) of peak-to-peak amplitude 10 μ m. The CMM, probe and thermal conditions are set to “perfect.” PUNDIT/CMM then predicts a cylindricity bias (mean error in measurement) of the (negative) full amplitude (-10 μ m) and no variability around this. To see why this is so, note that in simulation, the phasing of the sampling pattern to that of the form error is randomized on each cycle. But because the form error here is entirely systematic, and the number of sample points at each level is equal to the number of lobes, the cylindricity obtained in the simulated measurements is always *zero* regardless of the phasing. Thus on each cycle, the bias (measured cylindricity - true cylindricity) is the full (-) 10 μ m.

Whatever the phasing of the points to the lobing, the measured cylinder *center* location should remain fixed. And PUNDIT/CMM predicts zero bias and zero variability in measured position. Finally, the measured cylinder radius can be expected to range above and below its nominal value by half of the cylindricity, depending upon the phasing of the sampling to the lobing. The measured *diameter* would thus range over the full cylindricity (10 μ m). Furthermore, the distribution of the diameters can be expected to be quite non-gaussian, since the sinusoidal lobing form yields a notably higher sampling rate at its extrema than at intermediate values. This is just what PUNDIT/CMM shows in its error histogram.



Our second example illustrates the influence of CMM geometry on the uncertainty evaluation. Here we consider a CMM with a working volume of 500 x 500 x 600 mm and these B89 performance test parameters: linear accuracy (x,y,z) of 5.5, 6.5, 3 μ m respectively, volumetric performance 13.5 μ m, offset volumetric performance 50 ppm and repeatability 0.8 μ m. The workpiece is a ball step gage as shown. Each ball has a diameter of 25.4 mm and the step length is 100 mm. It is oriented parallel to the CMM z axis and placed near the xy center of the work volume. The probe, environment and workpiece form are taken to be perfect. All the spheres were measured with probing patterns that covered the available surfaces well. The spheres are numbered sequentially from top to bottom. The bottom sphere was taken as the coordinate system origin. The results of a run of 100 simulations are shown in the table below.

Sphere #	Uncertainties (μ m)			
	Z Location	XY Position	Sphericity	Diameter
1	10.4	22.7	5.0	2.1
2	8.4	18.8	4.9	2.5
3	6.4	15.0	4.7	2.5
4	5.0	11.6	4.7	2.4
5	4.5	9.2	4.8	2.5
6	2.2	5.6	4.8	2.5
7			5.1	2.9

The uncertainty of the z location increases with distance from the origin. It should be affected primarily by the z linear accuracy, the xz squareness and the yz squareness. Thus the primary B89 parameters that will govern this uncertainty will be the z linear accuracy and the volumetric performance. (Had an offset probe been used, the z axis roll would also have come into play via the offset volumetric performance.) At small distances from the origin, we would expect values on the order of the z linear accuracy; at larger distances we would expect values on the order of the volumetric performance. This is what we see in the table.

The uncertainty in the xy position should be primarily a function of xz and yz squareness (volumetric performance) and to some extent of x and y linear accuracy. At small distances from the origin the

uncertainty should be on the order of the linear accuracies; farther from the origin it should approach a value on the order of twice the volumetric performance. The data clearly reflect this trend as well.

Form and size uncertainties will be affected primarily by the three linear accuracy parameters and should be relatively independent of distance from the origin. This too is noted in the data.

Many tests of the sort described here, and others comparing PUNDIT/CMM results to specific CMM measurements on calibrated workpieces [7] have served to illustrate the validity of the methodology employed. PUNDIT/CMM can provide reliable, task-specific measurement uncertainties for GD&T parameters assessed by CMMs. Moreover, as a simulation system, it can do this off-line and even before workpiece production begins. Thus a viable measuring protocol can be established with confidence before the first article arrives for inspection. And by using NIST's Simulation by Constraints method, PUNDIT/CMM does not require sophisticated and time consuming studies of the CMM geometry.

Conclusion

We have described the design of a comprehensive system for evaluating the task-specific uncertainty of measurements made with coordinate measuring machines and shown examples of its application. With replacement of some of the CMM-specific models, it could be adapted to other 3D metrology systems. Application of this software will be beneficial in a variety of functions, including a) meeting new measurement traceability requirements, b) demonstrating product conformance to specifications, c) verifying that parts tolerances are complete, consistent and unambiguous, d) making the right choice when purchasing a CMM, e) finding and fixing the "weak link" in a CMM system, f) choosing the best CMM for a specific job, and g) training users in proper CMM measuring procedures.

References

1. International Standard 17025 General requirements for the competence of testing and calibration laboratories, (1999).
2. International Standard 14025-1 Geometrical Product Specifications (GPS)—Inspection by measurement of workpieces and measuring instruments—Part 1: Decision rules for proving conformance or nonconformance with specification. (1998).
3. ISO, Techniques of Determining the Uncertainty of Measurement in Coordinate Metrology, Technical Report 15530-1 (draft), July 16, 1998.
4. PUNDIT/CMM, © MetroSage, LLC, 2000-2004, <http://www.metrostage.com>
5. S. D. Phillips, B. R. Borchardt, D. Sawyer, W. T. Estler, D. Ward, K. Eberhardt, M. S. Levenson, M. McClain, B. Melvin, T. Hopp, Y. Shen, "The Calculation of CMM Measurement Uncertainty via The Method of Simulation by Constraints, *Proceedings of the 12th Annual Meeting of the American Society for Precision Engineering*, October 1997, Norfolk, VA, 443-446.
6. K. D. Summerhays, M. P. Henke, C. W. Brown, R. P. Henke, J. M. Baldwin, "A Tool for Determining Task-Specific Measurement Uncertainties in GD&T Parameters Obtained from Coordinate Measuring Machines," *Proceedings of the 17th Annual Meeting of the American Society for Precision Engineering*, Saint Louis, MO, October 22-24, 2002.
7. S. D. Phillips, B. Borchardt, A. J. Abackerli, C. Shakarji, D. Sawyer, P. Murray, B. Rasnick, K. D. Summerhays, J. M. Baldwin, R. P. Henke, M. P. Henke, "The Validation of CMM Task Specific Measurement Uncertainty Software," *Proceedings of the ASPE 2003 Summer Topical Meeting "Coordinate Measuring Machines"*, Charlotte, NC, June 25 -26, 2003.

STRUCTURAL INFERENCE FOR UNCERTAINTY EVALUATION WITH APPLICATION TO CALIBRATION EXPERIMENTS

Hari K. Iyer and C. M. Jack Wang
National Institute of Standards and Technology, Boulder, CO

In many situations, the value of a measurand M is related to certain quantities denoted by $m_1, m_2, \dots, m_N$ through a known functional relationship of the form

$$M = f(m_1, m_2, \dots, m_N).$$

The function f is generally a consequence of physical laws relating the relevant quantities. In practice, the values of m_i are not known exactly. However, results $X_1, X_2, \dots, X_N$ are available that estimate $m_1, m_2, \dots, m_N$, respectively. Since m_i are estimated based on direct measurements, we have a model of the form

$$X_i = m_i + b_i + e_i$$

where b_i are errors for which type-B information is available, and e_i are errors for which type-A information is available. Let Y be the estimate of M that is obtained from

$$Y = f(X_1, X_2, \dots, X_N).$$

Thus, the same functional relationship that relates the value M of the measurand to the input variables $m_1, m_2, \dots, m_N$ is used to calculate Y from the best estimates $X_1, X_2, \dots, X_N$ of $m_1, m_2, \dots, m_N$. Consequently, the error in Y must be calculated or estimated by propagating the errors in the quantities $X_1, X_2, \dots, X_N$ used to compute Y .

The ISO *Guide to the Expression of Uncertainty in Measurement* (GUM) recommends the use of a first-order Taylor series expansion for propagating errors and uncertainties. This generally provides adequate approximations for the actual uncertainties in the final result Y . However, when the nonlinearities in the function f are significant and/or the distribution of errors are described or modeled using a non-normal distribution, the GUM also permits the use of “other analytical or numerical methods” (Annex G.1.5). In this talk, we consider an alternative method, Fraser’s structural approach [1] [2], for the calculation of uncertainties in measurement results. The structural inference itself grew out of R. A. Fisher’s fiducial inference method. Consider a random sample $X_1, \dots, X_n$ of size n from a $N(\mu, \sigma^2)$ distribution. The unknown parameters of this model are μ and σ . Suppose that the parameter of interest is μ . It is known that $\bar{X}$ and S^2 are minimal sufficient and complete statistics for the parameters, where $\bar{X}$ is the mean and S^2 is the variance of the samples. Furthermore, it is known that $\bar{X}$ and S^2 are independent and have the following respective associated distribution

$$\begin{aligned} Z &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \\ U &= \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \end{aligned}$$

where $\chi^2(\nu)$ stands for the chi-squared distribution with ν degrees of freedom. We rewrite the above equations as follows:

$$\begin{aligned} \mu &= \bar{X} - \sigma Z/\sqrt{n} \\ \sigma &= \sqrt{(n-1)S^2/U}. \end{aligned}$$

In particular, substituting for σ in the first equation with the equivalent expression in the second equation, we get

$$\mu = \bar{X} - \left(\frac{S}{\sqrt{n}} \right) T_{n-1}$$

where $T_{n-1} = Z/\sqrt{U/(n-1)}$ is a student's t random variable with $n-1$ degrees of freedom. Fraser argued as follows. We have observed the value of $\bar{X}$ and S , say, $\bar{x}$ and s . Consequently, in the absence of any other information, all we know about μ is that it has a value that satisfies the equation

$$\mu = \bar{x} - \left(\frac{s}{\sqrt{n}} \right) T_{n-1}.$$

The only quantity that is unknown on the right hand side of the above equation is T_{n-1} . Thus, the value of μ must be related to a realization t_{n-1} of T_{n-1} via the above equation. Clearly, not all values of t_{n-1} are equally likely. Hence, not all values of μ are equally likely. The probabilities associated with the random variable T_{n-1} induce a probability structure on μ , and hence a distribution on μ . This induced distribution is what Fraser called the structural distribution of μ . We use $\tilde{\mu}$ as in

$$\tilde{\mu} = \bar{x} - \left(\frac{s}{\sqrt{n}} \right) T_{n-1}$$

to indicate the structural representation of the parameter μ . Inferences about μ can be made based on the structural distribution of μ . In this simple example, the structural procedures coincide with the classical frequentist procedures. For other complicated problems, the structural approach leads to solutions that are satisfactory and rivals other competing procedures.

To illustrate the use of the structural approach in calibration experiments, we consider the simple linear calibration model given by

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n$$

where Y_i are the observed responses, x_i are the standards, β_0 and β_1 are the intercept and the slope of the calibration line, and e_i are independent, normally distributed random variables with mean 0 and unknown variance σ^2 . We first assume that the standards x_i 's are known exactly. The unknown parameters for this model are β_0 , β_1 , and σ . The quantity of interest, for example, is the value of the standard x_0 when the observed response $Y = y_0$. That is,

$$y_0 = \beta_0 + \beta_1 x_0 + e_0$$

where e_0 is a normal random variable with mean 0 and standard deviation σ . The joint structural distributions of β_0 , β_1 , and σ , obtained from the sampling theory for the bivariate normal distribution, are given by

$$\begin{aligned} \tilde{\beta}_0 &= \hat{\beta}_0 - \frac{s}{\sqrt{U}} t_{11} Z_1 \\ \tilde{\beta}_1 &= \hat{\beta}_1 - \frac{s}{\sqrt{U}} (t_{21} Z_1 + t_{22} Z_2) \\ \tilde{\sigma} &= \frac{s}{\sqrt{U}} \end{aligned}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least-squares estimates of β_0 and β_1 , s^2 is the sum of squares of residuals from the fit, t_{ij} is the (i, j) th element of a lower triangular matrix t such that $(x^t x)^{-1} = t t^t$ with x being an $n \times 2$ matrix that the elements of the first column of the matrix all equal to 1 and the

second column contains x_i , Z_1 and Z_2 are independent standard normal random variables, and U is a χ^2 random variable with $n - 2$ degrees of freedom and is independent of Z_1 and Z_2 . Since

$$x_0 = \frac{y_0 - \beta_0 - e_0}{\beta_1} = \frac{y_0 - \beta_0 - \sigma Z_3}{\beta_1}$$

where Z_3 is a standard normal random variable independent of Z_1 , Z_2 and U . The structural distribution of x_0 can then be obtained as

$$\tilde{x}_0 = \frac{y_0 - \tilde{\beta}_0 - \tilde{\sigma} Z_3}{\tilde{\beta}_1}$$

and the statistical inference about x_0 may be made based on $\tilde{x}_0$.

If the values of standards x_i 's used in a calibration are not exactly known, but an estimate of x_i and its associated uncertainty are available, then the structural inference of this so-called *errors-in-variables* calibration model is just a straightforward extension of the inference of the simple calibration model. Specifically, the structural distributions of the unknown parameters of an errors-in-variables calibration model can be obtained by replacing x_i 's with their structural representations $\tilde{x}_i$'s in the calculation of $\hat{\beta}_0$, $\hat{\beta}_1$, s , and t .

In the context of uncertainty calculations, we can partition our measurement model parameters into two groups – parameters belonging to type-A distributions and those belonging to type-B distributions. For type-B parameters, the distributions that describe the uncertainties associated with the reported values of these parameters are assumed known. For type-A parameters, conditional on the values of the type-B parameters, we can now use their structural distributions to summarize the uncertainties associated with the reported values of these parameters. An important assumption that is made concerning the type-B parameters is that they are independent of the data. Hence the conditional distribution of the type-B parameters given the data is the same as their unconditional distribution. The net result is that we have a joint statistical distribution associated with the totality of all model parameters. It is now a straightforward technical matter to compute the distribution associated with any function of these model parameters. Such computations can be performed analytically in simple cases, or one can use simulation methods to estimate their distributions. For example, Let $Y_1, \dots, Y_n$ be n independent determinations of a measurand μ obtained from a single measurement experiment. They are assumed to follow the model

$$Y_i = \mu + b + e_i, \quad i = 1, \dots, n$$

where e_i and b are the type-A and type-B components of error. We also assume that e_i are independent normal random variables with zero mean and variance σ^2 , and b has a known distribution. Under the circumstances, it is shown that the structural distribution of μ is given by

$$\tilde{\mu} = \bar{y} - \frac{s}{\sqrt{n}} T_{n-1} - b$$

where $\bar{y}$ and s are the observed values of the sample mean and the sample standard deviation. The first two terms on the right hand side of the above equation account for the type-A component of the model, while the third term is for the type-B component. Since the distribution of b is completely specified, the distribution of $\tilde{\mu}$ may be determined analytically. However, it is most conveniently generated using Monte-Carlo simulation. A single realization of the structural distribution of μ may be generated as follows.

1. Generate a student's t random variables T_{n-1} .
2. Generate b according to its type-B distribution, independent of T_{n-1} .
3. Calculate $\tilde{\mu}$.

By generating a large number of independent realizations of $\tilde{\mu}$ we can estimate the mean, the uncertainty and quantiles of the structural distribution of μ and use them to make inference about μ .

Since the type-A and type-B components e and b may themselves be sums of elemental error components of their respective types, this simple example also demonstrates that the structural approach can be used to accommodate more general models for evaluating uncertainties in measurement results. It turns out that the use of structural distributions is particularly convenient especially when repeated measurements of an elemental quantity are assumed to follow a distribution which admits *complete sufficient statistics* as is the case in most metrological applications. More general cases can also be treated using the structural reasoning.

- [1] Fraser, D. A. S. *The Structure of Inference*, New York: Wiley, 1968.
- [2] Fraser, D. A. S. *Foundations of Statistical inference* (eds Godambe and Sprott), 32-55, Toronto: Holt, Rinehart and Winston, 1971.

Surface Figure Measurement at 80K: Alignment and Uncertainty Analysis

Peter Blake^{*a}, Ronald G. Mink^a, John Chambers^a, Pamela Davila^a,
David Content^a, F. David Robinson^{a,b},

^a Goddard Space Flight Center, Code 551, Greenbelt, MD 20771

^b Orbital Sciences Corporation, 7500 Greenway Center Drive, Suite 700, Greenbelt MD 20770

1. INTRODUCTION

The Optics Branch of the Goddard Space Flight Center (GSFC), interested in staying abreast of the rich, almost explosive development in new materials and forms for lightweight space-based astronomical mirrors, is developing its interferometry capabilities to new levels of precision. This paper presents our progress to date on the high-precision cryogenic facility. The facility is baselined for spherical mirrors with a radius of curvature (ROC) of 600 *mm*, and a clear aperture of 120 -- 150 *mm*. Our goal is to achieve uncertainties of 3 *nm rms* in the measurement at 20K of mirrors with a specified surface figure error (SFE) of 10 *nm rms*.

The near-term goals of the currently-reported experiment were:

- 1) Obtain the best possible estimates of the test mirror surface figure error (SFE) at both ambient conditions (room temperature) and at cryogenic temperatures. By SFE, we are referring to the two-dimensional map of deviation from the best-fit sphere; and we use the terms figure, surface figure, surface figure error, and error map synonymously.
- 2) Determine the uncertainty of these measurements.

2. TEST MIRROR AND TEST CONDITIONS

The first mirror tested was a flat-backed silicon foam-core mirror, concave spherical with radius of curvature of 600 *mm*, a thickness of 17.4 *mm*, and a clear aperture of 120 *mm*. The mirror was made by the Schafer Corporation in 2001, and is an early example of their SLMSTM technology¹.

To assure the highest possible accuracy and precision, it was decided that there would be no sources of mechanical strain introduced into the mirror by differential thermal contractions in the mount or in the heat strapping. The mirror would be simply-supported on two support points, with its flat back resting against the cold plate (fig. 1). There would be no constraint of the mirror and no thermal straps connected to the mirror.

The cold plate and inner shroud, which surround the mirror, except for the aperture, can be lowered to 12K with liquid helium. But with no heat straps and no thermal medium joining the mirror to the cold plate – just the three points of contact – the temperature performance of the mirror was unknown. Since radiative heat transfer scales as the fourth power of absolute temperature, once the mirror temperature falls below about 100K, the cooling of the mirror was expected to be slow, and would be countered by the small amount of radiation coming from the window at room temperature. As it turns out, the mirror reached a temperature of 95K with liquid nitrogen and 80K with liquid helium.

3. FACILITY

The test facility for cryo-measurement of the prototype mirrors was described in an earlier paper by these authors². A Zygo “Verifire AT”® phase measuring interferometer, positioned on crossed rails, focuses through a window into a cryostat or dewar. The cryostat has tip/tilt controls. The whole facility rests on a curtained vibration isolation table.

* peter.n.blake@nasa.gov; phone 301 286-4211; fax 301 286-0204

A simple redesign takes advantage of the converging interferometer beam for the testing of a 600 mm ROC sphere. We are able to use a small window and place it close to the focal point of the interferometer (figure 2) using a cylindrical extension that has been placed over the original window. There are several significant advantages: a) a small window can be thinner for a given deflection under vacuum, leading to less spherical aberration; b) a small footprint of the beam on the window makes it easier to achieve low transmitted wavefront error (WFE); and c) placing the window 600mm from the mirror on a narrow extension cylinder lowers the radiative coupling between the mirror and the window, lessening the mutual distortions of the window and the mirror.

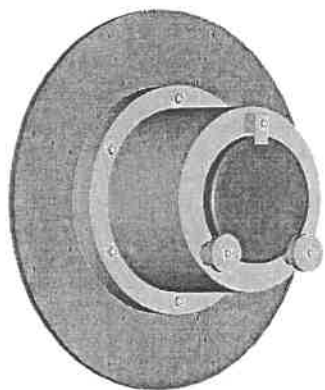


Fig 1: SLMS mirror supported on cold plate

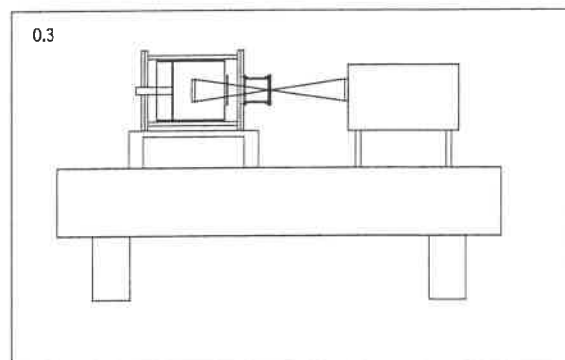


Fig 2: Dewar with window and snout added

4 CORRECTION OF SYSTEMATIC EFFECTS

As will be seen in the discussion of uncertainty analysis, we confined our search for and correction of systematic effects to those whose contribution to the SFE measurement was greater than 0.4 nm rms.

The primary systematic effect that requires correction is the aberration of the beam as it passes twice through the window of the cryostat. This aberration has been modeled using the software program Zemax®, as described in a previous paper³. It was found that if the optic axis were perpendicular to the window of the cryostat, the only aberration that contributed an SFE error of greater than 0.4 nm rms was spherical aberration, which contributed an error of 8.8 nm rms. It was found, however, that if the window were tilted, not only were there added contributions of coma and astigmatism, but the amount of aberration depended on the distance of the mirror from the window. Since this distance could change during thermal excursions, without our having any way to measure it, it was determined to keep the angle of the optic axis to the window less than 0.4 arcmin, giving an error of approximately 0.4 nm rms over and above the aberration at perpendicularity.

The window aberrations can be removed by two different calculations: 1) by subtracting the modeled or calculated aberrations from the final measured SFE; and 2) by measuring the change in SFE that occurs with the test mirror in the cryostat – behind the window – as the temperature drops from RT to cryogenic temperature (80K). When the measurement at room temperature and vacuum is subtracted from the measurement at 80K, the window aberrations of the two measurements – already shown by modeling to differ by less than 0.1 nm rms³ – are eliminated.

The method chosen should be the method with the least uncertainty. Our baseline approach, which is the only approach reported here, is the second: finding the change upon cooling (the “cryo-deformation” or the “cryo-difference map”) and adding this to the unwindowed room temperature (“ambient”) surface figure error.

In the interferometer itself, the only systematic effect above 0.4 nm rms is imperfection of the reference surface in the transmission sphere. This error can be measured and subtracted by a technique called “absolute measurement” or the “two-sphere test”^{3,4}. This technique will give both a corrected map of the test mirror SFE and a map of the interferometer errors over the pupil. This map of the interferometer error can be used as an error file to correct the measurements of test mirrors that do not lend themselves to the two-sphere test. For example, the ESA mirrors to be measured soon in our lab have integral feet for mounting, and would not be suitable for the two-sphere test, since gravity would force variations in surface figure as the mirror was rotated.

5 UNCERTAINTY DEFINED

Consider a result of an optical test W , which is an estimate of a surface figure. The result W is a set of N data points $\{w_1, w_2, \dots, w_N\}$ over a two-dimensional array. This measurement is the sum of the true surface figure error S , several contributions of systematic effects, and several contributions of statistical noise⁵. We will make efforts to remove the systematic effects, as described in section 4, but each correction has itself an unknown error, which must be estimated. The terminology becomes much clearer when we adopt the conception of *uncertainty*, as defined by NIST and ISO^{7,8}:

“Basic to the [ISO] approach is representing each component of uncertainty that contributes to the uncertainty of a measurement result by an estimated standard deviation, termed standard uncertainty, with suggested symbol u_i ...”

An uncertainty component, then, can be statistically determined – the statistically estimated standard deviations s_i , and the associated number of degrees of freedom ν_i ; or it may be approximated: “obtained from an assumed probability distribution based on all the available information.”

“The **combined standard uncertainty** of a measurement result, suggested symbol u_c , is taken to represent the estimated standard deviation of the result. It is obtained by combining the individual standard uncertainties u_i (and covariances as appropriate)...using...the *law of propagation of uncertainty*, the “root-sum-of-squares”...or “RSS” method of combining uncertainty components estimated as standard deviations....

It is assumed that a correction (or correction factor) is applied to compensate for each recognized systematic effect that significantly influences the measurement result and that every effort has been made to identify such effects. The relevant uncertainty to associate with each recognized systematic effect is then the standard uncertainty of the applied correction.”

Our task, then, is to estimate every individual uncertainty component, and sum them in quadrature.

5.1 Metrology goals

Our goal is to measure the cryo-figure of mirrors with specified SFE's of $< 10 \text{ nm rms}$. Since system level requirements for wavefront errors (WFE) are the sums in quadrature of the individual optics wavefront errors, and since it would be a reasonable goal to limit the uncertainty of the system-level WFE to 10%, the uncertainty goal for an individual measurement would be $\sqrt{0.1} \approx 0.3$, or 3 nm rms .

With an uncertainty goal of 3 nm rms , we limited our search to uncertainty components that were equal to or greater than 0.4 nm rms ; since ten such uncertainties, added in quadrature to other sources totalling 3 nm rms , would increase the total uncertainty by less than 10%.

5.2 Uncertainty as a root mean square

The root mean square (rms) of the surface deformation is a quantitative measure of the quality and performance of an optic, and, as such, is the final figure of merit in many investigations.

But how do we speak of the uncertainty of the surface figure itself – not the uncertainty of the rms figure, but the uncertainty of the two-dimensional map which is our estimate of the surface? The measurements, the corrections, and the final estimate of SFE are all maps over two dimensions. The uncertainties are not maps, *per se*, because they are unknown; but they are two-dimensional.

An approach has been suggested Ulf Griesmann, NIST⁶. The reported SFE is the sum of the true surface figure error S and unknown error. This error, although unknown, has an rms value. Our estimate of the expected range for rms value of the error is what we shall call the uncertainty of the measurement.

For example, following Griesmann, the short-term statistical component of uncertainty can be estimated by the following procedure:

- a Make multiple identical measurements (e.g., twenty times).
- b Average the multiple plots.
- c Subtract the mean plot – pixel by pixel -- from each measurement, to form what we'll call a delta.
- d Calculate the rms deviation of each of the deltas.
- e Determine the distribution of the rms values of the deltas: e.g., plot a histogram of frequency vs. rms; find the mean of the distribution and the standard deviation, s .
- f The uncertainty of the multiple measurements is defined such that 68% of the rms values are smaller than the uncertainty (68% confidence level analogous to the standard deviation). Thus uncertainty = mean + $0.468 * s$.

Because our interest is often in the surface figure itself, or in the power spectral density, we will carry through all computations of uncertainty as computations of the uncertainty of the surface figure, not uncertainty of the rms value itself, which can be calculated at the end.

6 MEASUREMENTS

Temperatures inside the dewar and in the test mirror were measured with temperature-sensing diodes. In order to limit the thermo-mechanical strain caused by attachment of diodes to the optic, the diodes were wired with 42 gage copper and attached to the mirror edge with a thin layer of GE Varnish.

When the cold plate was held at liquid nitrogen (77K), the mirror cooled down to 95K. When the cold plate was held at liquid helium temperature, the mirror cooled only to 80K. Our criterion for sufficient temperature stability for measurement was that the rate of change of the mirror fall below 1K/15 min.

The mirror was subjected to three thermal cycles, with measurements taken at the following points:

RTP: ambient conditions, room temperature, no window, no vacuum

RTV: Measurement taken through the window. Test mirror in dewar at room temperature under vacuum.

95K: Cold plate at liquid nitrogen temperature, mirror at 95K

80K: Cold plate at liquid helium temperature, mirror at 80K

When the rate of cooling was 1K/hr, the shutter was opened and the radiation from the window halted the cooling. On opening the shutter, the image is nulled and the measurement taken, as the mirror rises in temperature at close to 2K/hr.

For each measurement step of the thermal cycle, twenty successive individual measurements were taken, each using thirty-two phase averages.

7 RESULTS

In the cryo-cycling trials reported here, our correction for the error of the reference surface was to subtract the matching portion of the Zygo error file. The resultant SFE at 90% CA was found to be 22.6 *nm rms*.

Both systematic errors – window aberration and reference surface error -- are subtracted out when the data file for the mirror in the dewar, in vacuum at room temperature (RTVac or RTV) is subtracted from the data file for the mirror at 80K, after registering the files as closely as possible. The result was a cryo-change difference map with an rms of 4.9 *nm rms*. When the cryo-change difference map is added to the ambient condition measurement (corrected for reference surface error) the resultant SFE has an RMS of 25.0 *nm*.

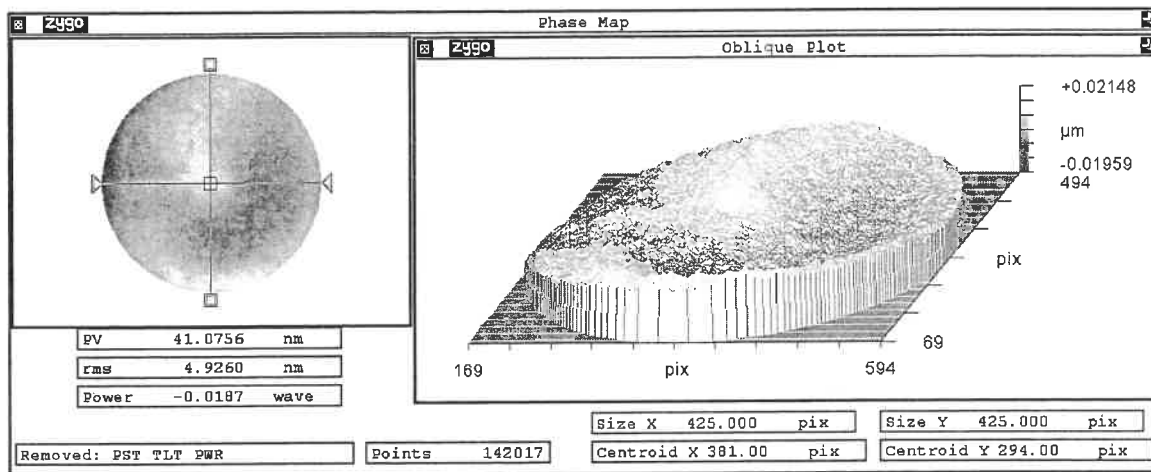


Figure 3: Change on Cooling of Schafer mirror

8 UNCERTAINTY ANALYSIS

8.1 Short-term statistical Uncertainty

When the twenty individual measurements are averaged, and the deviation of each measurement is subtracted from the mean, the short-term variation can be measured:

- at RTP, the short-term statistical uncertainty is: 0.7 nm rms
- at RTV, the short-term statistical uncertainty is: 0.9 nm rms
- at 95K, the short-term statistical uncertainty is: 0.8 nm rms
- at 80K, the short-term statistical uncertainty is: 1.2 nm rms

This is the uncertainty of any of the individual measurements. The mean of all twenty measurements for a condition would have lower uncertainty. The numbers quoted above include some OPD arithmetic uncertainty (see below); so an estimated short-term uncertainty of $\frac{1.2\text{nm}}{\sqrt{20}}$ turns out to be a reasonable estimate of the short-term statistical contribution.

8.2 Long-term statistical Uncertainty

A measure of the long-term statistical uncertainty was made in a previous experiment⁴, and was found to contribute (in RSS) an additional 0.8 nm rms to the statistical uncertainty.

8.3 Arithmetic Uncertainty

For each individual addition or subtraction of wavefronts (called OPD or *optical path length* arithmetic), a good estimate can be made of the uncertainty of position of the sets – the uncertainty of registration and congruence, pixel-for-pixel. The resultant error in the SFE stemming from that degree of misregistration can be quickly determined by taking an individual measurement, translating or mis-aligning it by that degree and subtracting it from the original. This estimation was done for every such addition or subtraction of data sets, with the calculated uncertainties ranging between 0.5 nm rms and 1.7 nm rms.

8.4 Optic Axis Alignment Uncertainty

Our technique for aligning the optic axis to the dewar window⁷ was accurate to within one pixel of the camera, which translates to an uncertainty of 0.4 nm rms.

8.5 Uncertainty of error file

To produce the error file used by us (4.4 *nm rms*), the Zygo corporation used a two-sphere test of the transmission sphere in our interferometer. We found it best to characterize this transmission sphere error by the first 36 Zernikes of the measured error file, losing the residual. This loss of the residual adds an additional uncertainty of 1.3 *nm*.

There is also uncertainty in the error file itself, which we estimated by examining the difference between two independent measurements: Zygo's measurement (using a convex mirror in a short cavity) and our own two-sphere measure of the transmission sphere, using the test mirror in the same cavity as is used in measuring the test mirror. The use of test mirror suffers the sole flaw of not filling the pupil. Over the common portion of the pupil, the two measurements differed by 1.4 *nm rms*; and that figure was taken as a conservative estimate of the uncertainty of the reference surface.

If when the two measurement files for 80K and RTV are subtracted, the two files do not occupy the same space in the pupil, there is small error contributed by the shear of the error file. This error is smaller than the OPD arithmetic necessary to eliminate it, so the uncertainty of the OPD arithmetic was conservatively assigned to this factor.

8.6 Combined Standard uncertainty

Combining these uncertainties in standard RSS form, with individual contributions from each iteration of OPD arithmetic, gives us an uncertainty in our map of cryo-deformation of 2.2 *nm rms*.

It is tempting to add and subtract in quadrature the estimated uncertainty of 2.2 *nm rms* with the estimated cryo-deformation of 4.9 *nm rms* in order to obtain an "error bar" between 4.4 *nm* and 5.4 *nm rms* for the surface figure RMS parameter. This does not appear to be legitimate, because several contributions to the uncertainty, such as the OPD arithmetic, are clearly not independent of the surface figure, and may be correlated with the cryo-deformation also.

More study of the propagation of uncertainty through these calculations is warranted.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of the Schafer Corporation, the European Space Agency, and the James Webb Space Telescope. Our thanks also to Chip Frohlich and Badri Shirgur of the Swales Corporation, Beltsville MD, and to Joe McMann of Mantech Systems Engineering Corp., Greenbelt, for their expert assistance on many aspects of this program. Special Thanks to Ulf Griesmann, NIST and David Content, GSFC.

Mention of trade names or commercial products does not constitute endorsement or recommendation by NASA.

REFERENCES

- ¹ Marc T. Jacoby, *et. al.*, "Helium cryo-testing of a SLMS™ (silicon lightweight mirrors) athermal optical assembly, *Proc SPIE 5180, Optical Manufacturing and Testing V*, ed. Phil Stahl, pp199-210, 2003
- ² P. Blake, R. G. Mink, D. Content, P. Davila, F. D. Robinson, S. R. Antonille, "Techniques and uncertainty analysis for interferometric surface figure error measurement of spherical mirrors at 20K", *Proc SPIE, 5180, Optical Manufacturing and Testing V*, ed. Phil Stahl, pp188-198, 2003.
- ³ J. E. Greivenkamp, J.H. Bruning, "Phase Shifting Interferometers", in *Optical Shop Testing*, 2nd ed., D. Malacara, ed., pp577-580, John Wiley & Sons, NY, 1992
- ⁴ K. Elssner, R. Burow, J. Grzanna, R. Spolaczyk, "Absolute Sphericity Measurement," *Applied Optics*, **28**, pp 4649-4661, 1989
- ⁵ A. Davies, M.S. Levenson, "Estimating the root mean square of a wave front and its uncertainty", *Applied Optics*, **40**, pp 6203-6209, 2001
- ⁶ Ulf Griesmann, National Institute of Standards and Technology, Gaithersburg, MD, personal communication.
- ⁷ To be published in *NASA Tech Briefs*

Radius of curvature uncertainty: Nonlinear measurand and treatment

Tony L. Schmitz^a and Angela Davies^b

^aDepartment of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611

^bDepartment of Physics and Optical Science, University of North Carolina at Charlotte, Charlotte, NC 28223

1.0 Introduction

The radius of curvature of an optical element is one of the dominant parameters that determines optical power and must be well-characterized if accurate results are to be obtained from ray trace models of complex systems. Radius errors can be partially compensated by element respacing during assembly, but this is time-consuming and costly and can lead to unwanted wavefront aberrations. For spherical components, radius is usually defined by the least-squares best-fit radius of the sphere over the clear aperture. Many measurement methods are available, including comparison with 'known radius' test plates, direct image formation/knife-edge test, astigmatism measurement, spherometers, autocollimator with pentaprism, and shearing interferometers, but the optical radius bench generally gives the lowest uncertainty [1-2].

Optical bench measurements are carried out using a figure measuring interferometer to identify two critical positions of the test artifact, confocal and cat's eye. The reflected wavefront matches the wavefront exiting the interferometer at these two locations, resulting in a nulled cavity. A linear transducer, often a displacement measuring interferometer or DMI, measures the displacement of the stage/part between the two positions and the recorded distance nominally represents the radius of the best-fit sphere to the surface. This measurement is shown schematically in Fig. 1 (after [3]), where the confocal and cat's eye positions are identified by a linear regression to the (circular aperture) Zernike polynomial α_2^0 , or power, terms recorded at a discrete number of locations along the optical, or z , axis of the interferometer (the corresponding DMI values are also recorded).

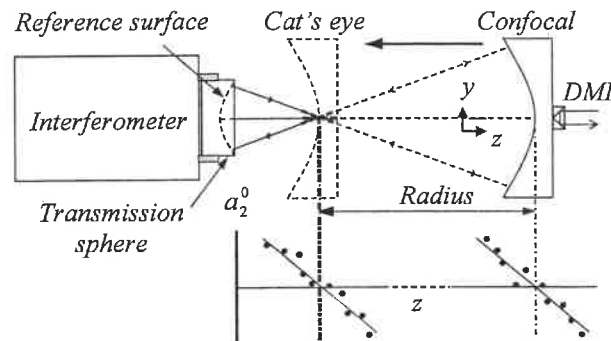


Figure 1: Schematic of optical bench measurement using a Fizeau interferometer.

To carry out a traceable radius measurement, however, all biases must be corrected and uncertainty sources combined to yield the combined standard uncertainty, u_c , which represents one standard deviation of the measurand, or quantity under test [3-7]. In an effort to properly define the measurand for optical bench measurements, we have implemented a homogeneous transformation matrix (HTM) formalism, where a vector equation is used to define the radius and an HTM is used to determine the location of the test artifact after translation from confocal to cat's eye. It is shown that for the nonlinear measurand equation generated by this approach, it is not possible to simply insert expectation values for each uncertainty contributor to determine the best estimate of the measurement variance. Rather, a Monte Carlo simulation is applied to determine the expected measurement variance.

2.0 Radius equation description

There are many potential sources of radius uncertainty and measurement bias, such as wavefront aberrations, null identification at confocal and cat's eye, artifact figure error, displacement gage error, and stage error motions [3, 6-7]. Error motions are particularly challenging because they are interrelated and convolved with the measurement. Recent work has focused on identifying uncertainty contributions, removing measurement biases, and determining the uncertainty through a root-sum-of-the-squares (RSS) combination [3, 7]. However, an equation that explicitly defines radius as a function of all uncertainty sources was not implemented. In the HTM approach presented here, the measurand, R , is represented by a vector equation that intrinsically corrects for error motions and allows other uncertainty contributors to be conveniently represented.

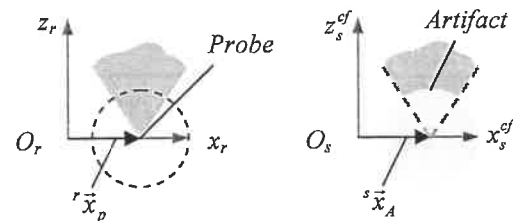


Figure 2: Coordinate frame identification for the confocal position. Reference and stage frames overlap at this location.

Formulation of the HTM-based radius equation begins with identification of the necessary coordinates and frames (similar to kinematic error analyses for precision machine tools, robots, and stages [8]). These include the reference, r , and stage, s , frames and their origins, O_r and O_s , respectively, and the probe, p , and artifact, A , coordinates, where p is assumed to be located at the geometric focus of the interferometer and A at the artifact best-fit center. Figure 2 shows the frames and origins for the confocal position, as well as the vectors ${}^r\bar{x}_p$ and ${}^s\bar{x}_A$, which locate the probe in the reference frame and artifact center in the stage frame, respectively. Two measurement scenarios are depicted in Fig. 3.

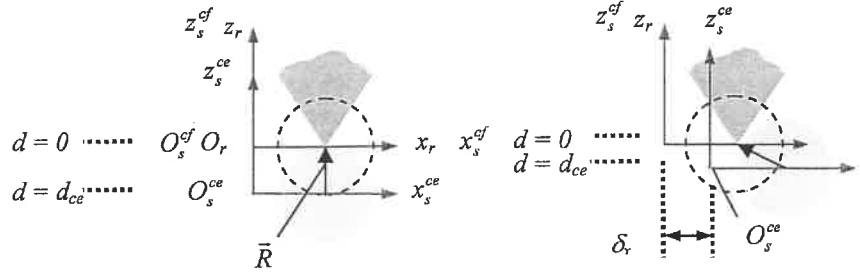


Figure 3: (Left) error free radius measurement; (right) straightness error introduces measurement bias.

For the left figure, there is perfect translation between confocal and cat's eye along the z_r axis. In this case, the displacement gage reading at cat's eye, d_{ce} , is equal to the radius amplitude, R . In the right figure, an x -direction straightness error, δ_x , is imposed and the gage reading is now less than the true radius (i.e., a measurement bias is introduced).

The straightness error scenario is reproduced in Fig. 4 to describe the vector radius equation. The figure shows that the radius can be expressed as the vector difference, $\bar{R} = {}^r\bar{x}_p - {}^r\bar{x}_A^{ce}$, and the amplitude by $R = |\bar{R}| = \sqrt{{}^r\bar{x}_p - {}^r\bar{x}_A^{ce}{}^2}$.

To solve these equations, it is necessary to express ${}^r\bar{x}_p$ and ${}^r\bar{x}_A^{ce}$ in terms of measurable quantities. We can express the latter using an HTM, populated by separate measurements of the stage error motions, to map the artifact center's location at cat's eye to the reference coordinate system. The required HTM is the following:

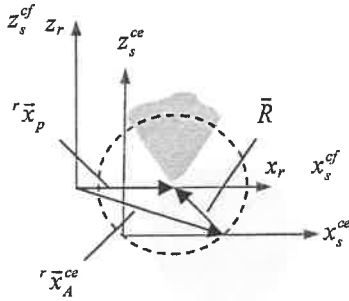


Figure 4: Vector radius description.

during z -direction motion of the stage (relative to the reference frame) and/or 2) misalignment between the gage axis and reference frame (i.e., traditional cosine error). We can then write ${}^r\bar{x}_A^{ce} = {}^rT_s {}^s\bar{x}_A$. In this equation, the artifact center in the stage frame, ${}^s\bar{x}_A$, may be written as a function of probe offsets in the reference frame and the ability to null at the confocal position, dx^{cf} , dy^{cf} , and dz^{cf} (note that the reference and stage frames ideally overlap at confocal). After these substitutions, the final vector equation can be written as shown in Eq. 2, where R^2 is determined from the sum-of-the-squares of the x , y , and z -direction components of the right hand side.

$${}^rT_s = \begin{bmatrix} 1 & -\varepsilon_z & \varepsilon_y & \delta_x \\ \varepsilon_z & 1 & -\varepsilon_x & \delta_y \\ -\varepsilon_y & \varepsilon_x & 1 & \delta_z + d \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

$$\bar{R} = {}^r\bar{x}_p - {}^rT_s \begin{bmatrix} dx^{cf} \\ dy^{cf} \\ dz^{cf} \\ 1 \end{bmatrix} = \begin{bmatrix} -dx^{cf} + \varepsilon_z({}^r y_p + dy^{cf}) - \varepsilon_y({}^r z_p + dz^{cf}) - \delta_x + 0 \\ -dy^{cf} - \varepsilon_z({}^r x_p + dx^{cf}) + \varepsilon_x({}^r z_p + dz^{cf}) - \delta_y + 0 \\ -dz^{cf} + \varepsilon_y({}^r x_p + dx^{cf}) - \varepsilon_x({}^r y_p + dy^{cf}) - \delta_z - d \end{bmatrix} \quad (2)$$

I
II
III
IV

2.1 Analytical expectation value

Four error categories are identified in Eq. 2. The translational errors in column I represent the ability to null at confocal. Category II groups the Abbe errors (rotation-offset products), while category III gives the straightness/cosine errors. The fourth column consists of the gage reading for z-direction motions. Equation 2 is expanded in Eq. 3 to give the final R expression. To simplify this expression, we have set the terms containing δ_x , which represents positioning error along the reference axis, and r_{z_p} , the probe offset in the z-direction expressed in the reference frame, equal to zero. The positioning error along the z-axis is accounted for by calibrating the displacement gauge with some uncertainty. We assume the error motions depend only linearly on the z-position. Therefore, the location of $z = 0$ is arbitrary and r_{z_p} can be set to zero.

$$R = \left(\begin{aligned} & d_{ce}^2 + \delta_x^2 + \delta_y^2 + dx^{ef2} + dy^{ef2} + dz^{ef2} + 2dx^{ef}\delta_x + 2dy^{ef}\delta_y + 2dz^{ef}d_{ce} + \\ & \varepsilon_z^2(r_{y_p}^2 + 2r_{y_p}dy^{ef} + dy^{ef2}) + \varepsilon_z^2(r_{x_p}^2 + 2r_{x_p}dx^{ef} + dx^{ef2}) + \\ & \varepsilon_y^2(r_{x_p}^2 + 2r_{x_p}dx^{ef} + dx^{ef2}) + \varepsilon_x^2(r_{y_p}^2 + 2r_{y_p}dy^{ef} + dy^{ef2}) + \varepsilon_x^2(dz^{ef2}) + \varepsilon_y^2(dz^{ef2}) + \\ & 2\varepsilon_z\delta_y r_{x_p} - 2\varepsilon_z\delta_x r_{y_p} + 2\varepsilon_x r_{y_p}d_{ce} - 2\varepsilon_y r_{x_p}d_{ce} + 2dy^{ef}\varepsilon_x d_{ce} - 2dx^{ef}\varepsilon_y d_{ce} + \\ & 2dz^{ef}\varepsilon_z(r_{x_p} - \delta_x) - 2dx^{ef}\varepsilon_z(r_{y_p} - \delta_y) + 2dz^{ef}\varepsilon_x(r_{y_p} - \delta_y) - 2dz^{ef}\varepsilon_y(r_{x_p} - \delta_x) - \\ & 2\varepsilon_x\varepsilon_y(r_{x_p}r_{y_p} + r_{x_p}dy^{ef} + r_{y_p}dx^{ef} + dx^{ef}dy^{ef}) - 2\varepsilon_x\varepsilon_z(r_{x_p}dz^{ef} + dx^{ef}dz^{ef}) - 2\varepsilon_y\varepsilon_z(r_{y_p}dz^{ef} + dy^{ef}dz^{ef}) \end{aligned} \right)^{\frac{1}{2}} \quad (3)$$

Given this expression for R , a logical next step would be to determine the best estimate of a measured radius by evaluating the equation for each term on the right hand side by using the expectation value, which is defined as the product of the parameter and its probability distribution integrated over all possible values [9]. Although it may not be the case in general, for the discussion here we will assume here that all errors are zero on average and the gage axis is collinear with the motion axis (recall that a misalignment will introduce a linear term into the straightness values). The expectation value of each non-squared errors is, therefore, zero. By applying this assumption and neglecting terms that are products of higher order (these terms would be small for a well-designed system), Eq. 3 can be approximated as shown in Eq. 4. The reader may note that the squared terms remain because the expectation values are not zero, but rather are equal to the variance (or square of the standard deviation).

$$R \approx \left(\begin{aligned} & d_{ce}^2 + \delta_x^2 + \delta_y^2 + dx^{ef2} + dy^{ef2} + dz^{ef2} + \\ & \varepsilon_z^2(r_{y_p}^2 + dy^{ef2}) + \varepsilon_z^2(r_{x_p}^2 + dx^{ef2}) + \\ & \varepsilon_y^2(r_{x_p}^2 + dx^{ef2}) + \varepsilon_x^2(r_{y_p}^2 + dy^{ef2}) + \varepsilon_x^2(dz^{ef2}) + \varepsilon_y^2(dz^{ef2}) \end{aligned} \right)^{\frac{1}{2}} \quad (4)$$

To determine the expectation value for R , common practice would be to substitute the expectation value for each input on the right hand side of Eq. 4. In fact, some of the remaining terms make this an appealing option. For example, the presence of the straightness error, δ_x , would account for the bias shown in Fig. 3 which causes the gage reading to be less than the actual radius amplitude. The reported measurement value, R , is compensated for this error since it is summed in quadrature with d_{ce} (which would otherwise be used as the reported radius value). However, due to the nonlinear nature of the radius equation, this direct term-by-term expectation value substitution cannot be applied. The error is revealed if we consider the dz^{ef2} term. Recall that dz^{ef} is the ability to null the confocal position at the beginning of the optical bench measurements. Let's assume that all other error contributions are exactly zero. If this z-direction error is taken to have a mean zero and normal distribution with standard deviation σ_{dz} , we would expect the measured radius to be equal to the mean gage reading (recorded along the z axis) with a normal distribution and standard deviation σ_{dz} . However, Eq. 4 suggests that a bias in R dependent on σ_{dz}^2 would be introduced. This is clearly not the case and highlights the invalidity of this approach. Said another way, for a generic

parameter x , to which is added noise δx with a mean of 0, the best estimate for x is $x = \left\langle \sqrt{(x + \delta x)^2} \right\rangle$ which is *not* equal to $\sqrt{\left\langle (x + \delta x)^2 \right\rangle}$ (where the angled brackets represent the expectation value).

2.2 Monte Carlo analysis

For situations like this where the measurand equation is complicated and nonlinear, we recommend a Monte Carlo simulation to directly evaluate the mean and standard deviation in R given the input distributions and mean values of the inputs to Eq. 3 (no truncation of the original equation need be applied). Estimates of the expected values, uncertainties, and the form of the probability distribution (usually Gaussian) for each term in Eq. 3 are needed to carry out the simulation and estimate R for a particular measurement. A Monte Carlo analysis shows that the category I and III terms in Eq. 2 give single-sided (biased) distributions in R for the x and y -direction components, but unbiased results for the z terms. See Figs. 5 and 6, where 100,000 points and zero probe offsets were applied in the simulations. These results agree with the intuitive description provided in Section 2.1 and verify that a direct analytic substitution into the nonlinear Eq. 3 yields incorrect results.

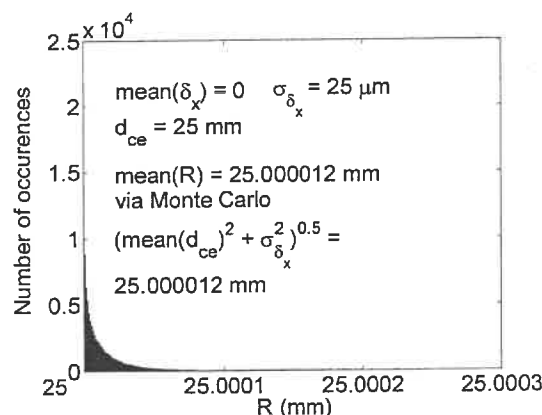


Figure 5: Histogram of Monte Carlo results for mean zero δ_x error with standard deviation of $25 \mu\text{m}$. A single-sided (biased) distribution is observed.

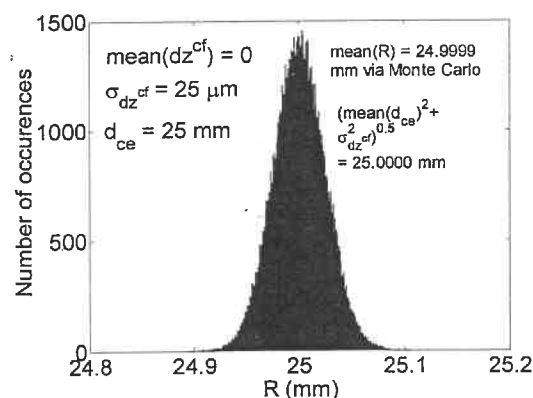


Figure 6: Histogram for dz^{cf} error with zero mean and $25 \mu\text{m}$ standard deviation. No bias is introduced in this case.

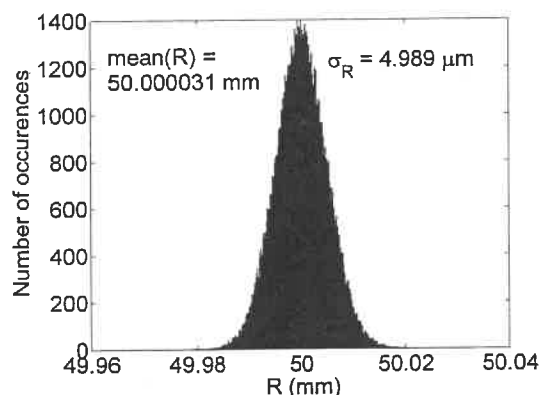


Figure 7: Monte Carlo simulation results for measurement described in Table 1.

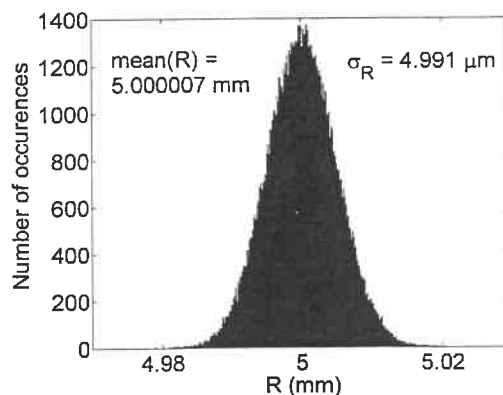


Figure 8: Monte Carlo results for $d_{ce} = 5 \text{ mm}$.

2.3 Uncertainty evaluation

Two options may be considered to determine $u_c(R)$ from Eq. 3. Following the ISO GUM [4] recommendations, the Taylor series expansion of R can be completed and the standard uncertainties in each variable inserted into the

result and multiplied by the corresponding sensitivity coefficient (i.e., the partial derivative of R with respect to the input variable in question). This would be extremely difficult. Instead, the distribution of R values that results from the Monte Carlo simulation directly reflects the combined standard uncertainty. In other words, by defining the mean, standard deviation, and distribution of each input to Eq. 3 using Type A (statistical) or Type B (other) approaches [4-5], the resulting standard deviation from the Monte Carlo simulation is u_c . As an example, the distribution of R values shown in Fig. 7 is the simulation result for the normally distributed inputs shown in Table 1. Although x and y -direction straightness and confocal null errors are present, this distribution is not single-sided. This is due to the nonzero z error standard deviations. The standard deviation of the 100,000 Monte Carlo simulation data points is $5.0\text{ }\mu\text{m}$; therefore, $u_c(R) = 5.0\text{ }\mu\text{m}$ (0.01% of the mean) for the conditions given in Table 1. As a comparison, this was repeated for $d_{ce} = 5\text{ mm}$ using the same input uncertainties (this is reasonable since the uncertainties are based on our ability to measure the individual uncertainty contributors and not always directly on the size of the artifact in question) – see Fig. 8. In this case, the mean R was 5.000007 mm and $u_c(R)$ was $5.0\text{ }\mu\text{m}$ (0.1%). This underscores the importance of improved optical bench characterization and/or precision construction for measurement of small radii.

Table 1: Input values for Monte Carlo simulation.

Input	Mean	Standard deviation
r_{x_p}	2 mm	50 μm
r_{y_p}	2 mm	50 μm
r_{z_p}	0	50 μm
d_{ce}	50 mm	50 nm
dx^{cf}	0	1 μm
dy^{cf}	0	1 μm
dz^{cf}	0	5 μm
δ_x	0	1 μm
δ_y	0	1 μm
δ_z	0	0
ε_x	0	5 μrad
ε_y	0	5 μrad
ε_z	0	5 μrad

3.0 Conclusions

In this paper, we described an HTM-based vector equation which defines the measurand in optical bench radius measurements for spherical optics. It was shown that for the corresponding nonlinear radius equation, direct substitution in order to analytically determine the expectation value leads to erroneous results. Additionally, the use of Taylor series expansion of the measurand to determine the combined standard uncertainty is not recommended. As an alternative, Monte Carlo simulations are successfully applied to conveniently find both the mean and standard deviation, which represents the combined standard uncertainty, of the radius. This was illustrated for multiple cases.

4.0 Acknowledgements

The authors wish to acknowledge helpful discussions with W. T. Ester, National Institute of Standards and Technology, Gaithersburg, MD, and C. J. Evans, Zygo

Corporation, Middlefield, CT, during this work.

5.0 References

1. Malacara, D., 1992, *Optical Shop Testing*, John Wiley & Sons, New York.
2. Murty, M. and Shukla, R., 1983, Measurements of long radius of curvature, *Optical Engineering*, **22**: 231-235.
3. Schmitz, T. L., Evans, C. J., Davies, A., and Estler, W. T., 2002, Displacement uncertainty in interferometric radius measurements, *Annals of the CIRP*, **51/1**: 451-454.
4. International Organization for Standardization (ISO), 1995, *Guide to the Expression of Uncertainty in Measurement*, Geneva.
5. Taylor, B. and Kuyatt, C., 1994, NIST Technical Note 1297, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*.
6. Selberg, L., 1992, Radius measurement by interferometry, *Optical Engineering*, **31**: 1961-1966.
7. Schmitz, T. L., Davies, A., and Evans, C. J., 2001, Uncertainties in interferometric measurements of radius of curvature, *Optical Manufacturing and Testing IV*, Stahl, H. P., Editor, *Proceedings of SPIE*, **4451**: 432-447.
8. Slocum, A., 1992, *Precision Machine Design*, Prentice-Hall, Englewood Cliffs, NJ.
9. Bevington, P. R. and Robinson, D. K., 1992, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd Ed., WCB/McGraw-Hill, New York.

ACCURACY EVALUATION OF WHITE LIGHT INTERFEROMETER REFERENCE SIGNAL

Der-Shen Wan, Eric Frey and Erik. Novak
Veeco Instruments, 2650 E. Elvira Road, Tucson, AZ 85706

White light interferometry is a powerful tool for surface profiling over large vertical ranges (up to 8 mm). Using the principle of independently determining the zero optical path difference (OPD) condition for each point within the field of view, the profile of the test surface is determined without the step and slope limitations associated with phase shifting interferometry. The zero OPD condition can be determined with accuracy on the order of a small fraction of the wavelength of light used ($1/100^{\text{th}}$ to $1/1000^{\text{th}}$ of wavelength). Various methods of finding the zero OPD position have been implemented to achieve the accuracies mentioned above. These include the use of maximum fringe visibility, zero-phase fringe location, correlation methods, wavelets, and other techniques [1][2]. This method can be implemented on a microscope-based profiler to provide high vertical and lateral resolution. High accuracy and non-contact operation has made these systems widely employed as the metrology tool of choice for such fields as MEMS, semiconductors, data storage, micro-optics, and medical instrumentation.

The way to achieve the zero OPD condition using an interferometric profiler is to scan either the optical profiler or the test surface with respect to the other, until the OPD has been varied over the entire height range of the test piece. When a feature on the test surface is well focused, the location of zero OPD is found by the methods stated above. The accuracy of a profile measurement using any of the zero-OPD determination techniques is therefore highly dependent on knowing the exact behavior of the scan mechanism. Scanners have been calibrated using known steps with heights similar to those of the test objects. The step standards, however, have higher uncertainties than many processes now require. The validity of the step calibration is also subject to change due to wear, friction and other factors, which can reduce the repeatability of the scanner motion. In order to improve the knowledge of the scanner position, to the accuracy required by the most challenging applications, we have recently implemented a HeNe laser distance measurement interferometer (DMI) on an optical profiler. The DMI continuously monitors the scanner location and this information is used to dynamically correct the algorithms involved [3]. This distance measurement information is called 'reference signal for white light interferometry'. Figure 1 shows schematically an optical profiler with a reference signal. In order to evaluate the accuracy of this reference signal, we will analyze the various sources of error and use experiments to find the dominant error on a reference signal prototype. Even with these errors, this prototype profiler with the reference signal shows an order of magnitude improvement in accuracy over those calibrated with the step height method.

As in any distance measurement, environmental effects need to be considered first. The distance measured by the reference signal is the optical path, which is the distance times the refractive index of air. The change of refractive index of air with temperature, pressure and relative humidity are calculated according to a general formula [4]. For

normal lab environments, as temperature is maintained between 24°C to 25°C, pressure from 29.9 to 30 in Hg, and humidity variation within 10%, the change of optical path due to these variations over a distance measurement of 1 mm are: 1 nm, 1 nm and 0.1 nm respectively. This is less than other errors associated with DMI itself.

We will discuss three other sources of error associated with this reference signal: cosine error, Abbe error and the phase fitting or non-linearity error.

Cosine error results from non-parallelism between the reference signal DMI laser beam and the axis of scanner motion. This makes the measured distance shorter than actual as expressed by $L \cos(\alpha)$, where L is the actual distance and α is the angle between the laser beam and the axis of scanner motion. Since α is typically very small, the difference between L and $L \cos(\alpha)$ can be expressed as $L\alpha^2/2$. In the prototype system, the highest contribution to this error is from mechanical tolerance stack-up and it is estimated to be 0.3°. The calculated cosine error for a 1 mm distance measurement is 27 nm. This error can be decreased by more careful control of mechanical tolerances.

Abbe error is a factor in the prototype since the reference signal axis is offset from the OPD or scanner measurement axis. Angular irregularities (tip and tilt) in scanner stage motion cause axial motion at off-axis points equal to the sine of the rotation angle multiplied by the distance from the center of rotation. Figure 2 diagrams Abbe error from a scanner stage rotation. The reference signal and the OPD axes are both offset from the rotation center and not coincident so the difference between their Abbe error induced axial motions will be equivalent to the sine of the angle multiplied by the distance between the axes.

In order to quantify the Abbe error, two DMIs and a Michelson interferometer were used on the scanner stage, as shown in Figure 3. One DMI is placed at the designed position for the reference signal in the profiler prototype (25mm from the optical axis). The second DMI is placed on the optical axis and directly evaluates motion of the microscope objective. The Michelson interferometer is set to measure the tip and tilt of the scanner stage as it translates. The tip and tilt information is used to determine if the difference between the two DMI measurements is from Abbe error. Figure 4(a) shows typical results of this setup for the difference between the two axes measured. The peak to valley difference for a scan distance of 1 mm is 0.45 micron. The predicted Abbe error, calculated using the product of the axis spacing and the measured tip and tilt angle, is shown in Figure 4(b). The calculated peak to valley change for a scan distance of 1mm is 0.35 micron. Comparison of the two shows good correlation and indicates the Abbe error for this prototype system is the dominant factor in the difference between the two axes. The discrepancy between the difference result and the tip and tilt result is probably caused by uncertainty in the tip and tilt measurement. The large scaling factor, due to spacing between the axes, means that 0.8 arc-seconds of tip and tilt corresponds to 0.1 microns of motion difference along the translation axis. The repeatability of the tip and tilt measurement on a stationary scan stage over ten minutes is 0.6 arc-seconds. This shows that Abbe error is a dominant factor so a modified optical design, placing the DMI on the same axis as the objective, could greatly reduce the overall system error.

Finally we tested the phase fitting algorithm for the reference signal distance measurement. Here the distance measurement is based on quadrature detection [5], and the Lissajous figure from the outputs of two detectors is shown in Figure 5. As the scanner stage moves a longer distance the Lissajous figure traces over itself at slightly different position due to the non-linearity of the scanner. The distance is found by fitting ellipses for the Lissajous figures. In order to check the validity of the fitting algorithm, we compared the distance retrieved from fitting the Lissajous figure shown in Figure 6 just once with the distance found by fitting fifty consecutive Lissajous figures. The results of the one-time fitting and the fifty-times fitting are within 1 nm. This shows that the ellipse-fitting algorithm is robust and able to detect the non-linearity of the scanner.

In conclusion, step height calibration based optical profiler system uncertainty of 0.5% can be improved to 0.05% (0.4 micron / 1 mm) by implementing the HeNe laser based DMI. Residual errors, primarily Abbe but also cosine error still exist but can be greatly reduced by careful redesign of the mechanics. These changes could potentially lead to accuracy in the range of $1/100^{\text{th}}$ to $1/1000^{\text{th}}$ wavelength over an 8 mm scan length. That will be another one to two order of magnitude reduction of uncertainty from the current prototype.

References:

1. Harasaki A., J. Schmit and J.C. Wyant, *Appl. Opt.*, 39, 2107 (2000).
2. De Groot P., and L. Deck, *J. Mod. Opt.*, 42, 389 (1995).
3. Olszak A. G., and J. Schmit, *Opt. Eng.*, 42, 54 (2003).
4. Owens J.C. *Appl. Opt.*, 6, 51, 1967.
5. Heydemann P., *Appl. Opt.*, 20, 3382 (1983).

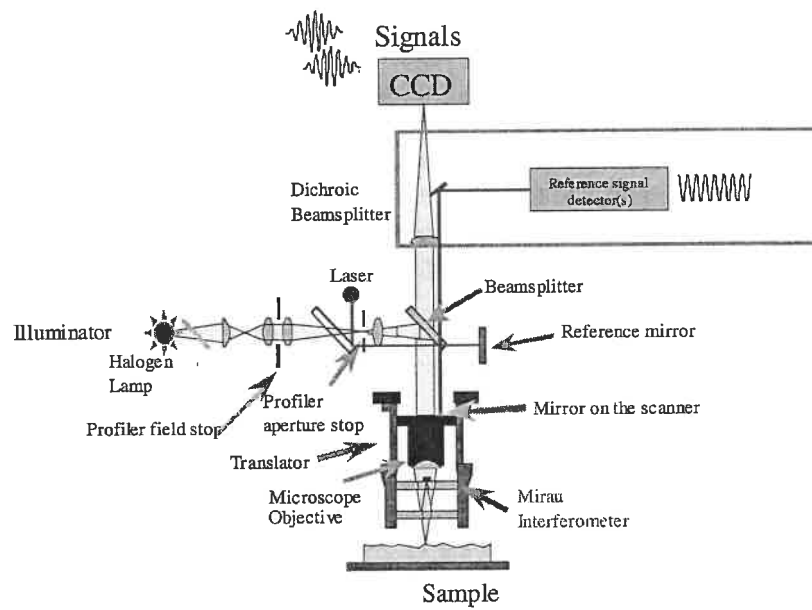


Fig. 1 Schematic diagram of optical profiler with reference signal.

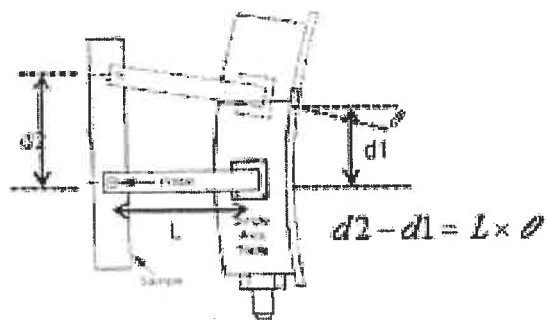


Fig. 2 Abbe error.

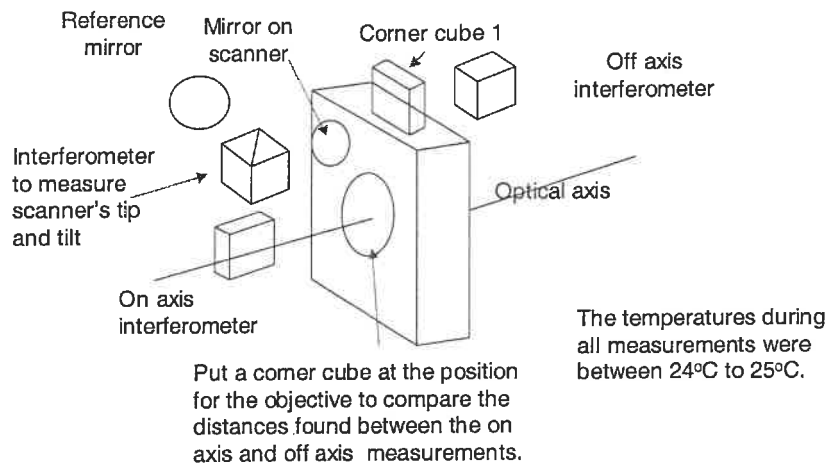
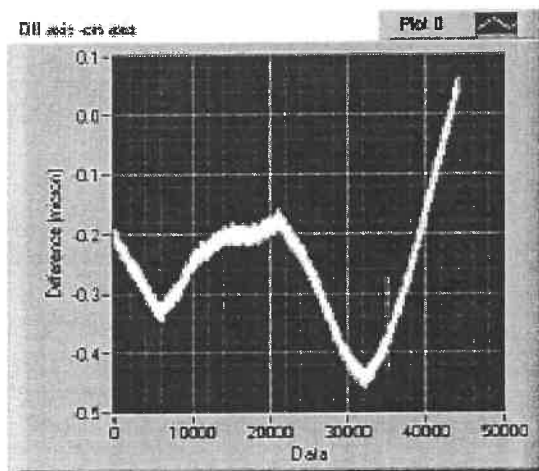
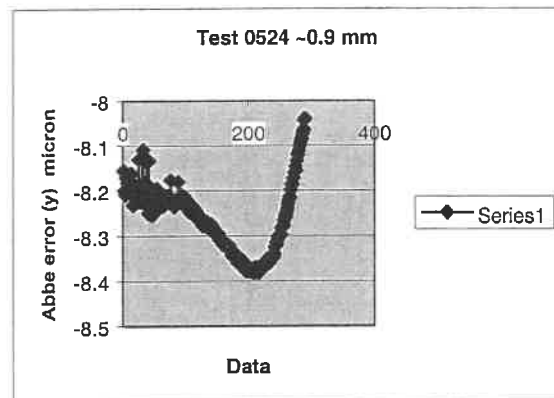


Fig. 3 Experiments to measure Abbe error.



(a) Difference between off axis and on axis measurement over 0.96 mm. Peak to valley 0.45 micron.



(b) Abbe error found by tip and tilt measurement. Peak to valley 0.35 micron.

Fig. 4 Abbe error measurement results.

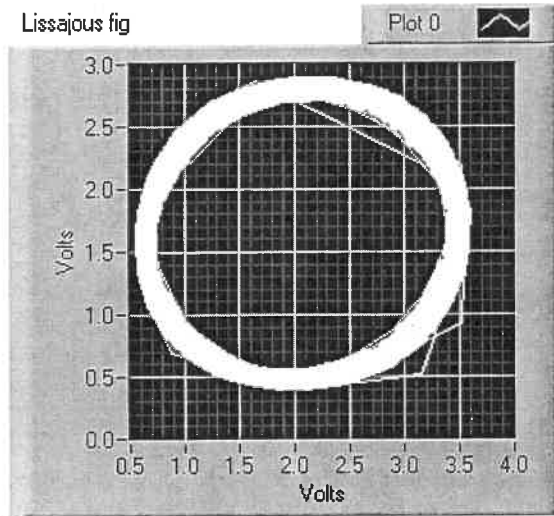


Fig. 5 Lissajous figures of scanning over 1 mm.

DISPLACEMENT MEASURING INTERFEROMETRY MEASUREMENT UNCERTAINTY

Mike Holmes, Chris Evans
Zygo Corporation, Middlefield, CT 06455

Introduction

Displacement measuring interferometry is a key enabling technology allowing the semiconductor industry to continue to shrink linewidths by a factor of two every 2-3 years and increase integrated circuit performance. Intel, for example, have indicated that they expect to see limited production at the 65 nm node in 2005; line widths down to 32 nm by the end of the decade are predicted in the most recent ITRS roadmap. Typically, overlay budgets are about 1/3 the linewidth and reticle and wafer stage metrology themselves are but one of many factors contributing to overlay budget. Hence it is clear that sub-nm wafer stage position uncertainties will be needed on the stages in the lithography tools used to make the integrated circuits.

Nanometer class position uncertainties are required in other applications, such as a long travel measuring system being developed at Zygo. In this application, the position of a 6 degree of freedom stage with a working volume of 500 x 1 x 1 mm must be known with an uncertainty of 1 nm ($k=1$). Including calibration tools and the wavelength tracker, 14 DMIs are used in four different configurations. This presentation focuses on one of those configurations, and gives a detailed analysis of the uncertainties; the methodology is shown here for two uncertainty contributors.

Ideal

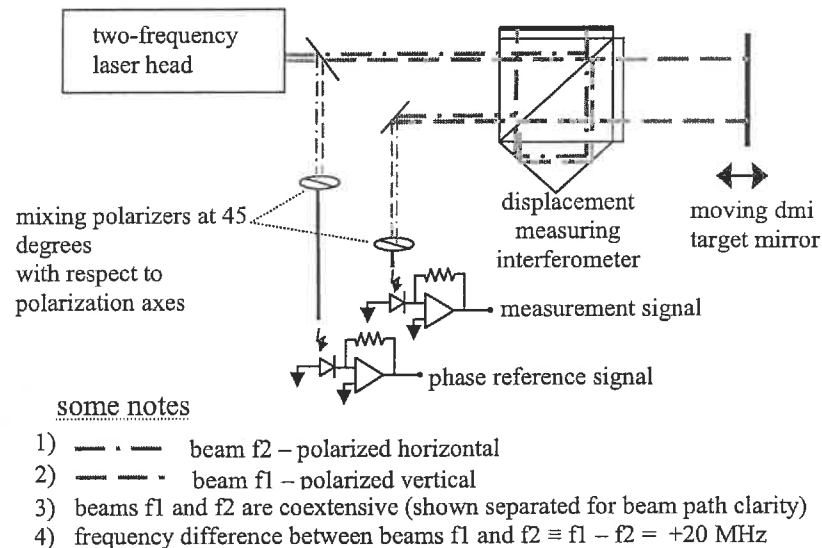


Figure 1 – Ideal heterodyne displacement measuring interferometer system minus phase interpolator

Figure 1 shows an idealized heterodyne Displacement Measuring Interferometer (DMI) system^{1,2} consisting of the laser head, displacement measuring interferometer, moving target mirror, and transimpedance amplifiers. The output from the laser head is two coextensive linearly polarized beams ($\lambda_{1,2} \approx 632.991$ nm) which are separated in frequency by 20 MHz. The polarization states are orthogonal to each other. The two components are separated in the interferometer. One component travels a constant optical path through the reference arm of the interferometer while the other travels through the measurement arm whose optical path length is dependent upon the position of the target mirror. The two components rejoin inside the interferometer and are mixed and directed onto the measurement signal transimpedance amplifier. The electrical output from this amplifier is a $20 \text{ MHz} \pm \text{Doppler shift}$ (a Doppler shift occurs when the target mirror is moving and is proportional to its velocity) sinusoid. This signal's magnitude is

proportional to the heterodyne signal optical power and its phase is proportional to the interferometer's measurement arm optical path length. The electrical phase reference signal is generated from a portion of the interferometer's input beams. Note that this signal has constant phase as the two components are traveling a common constant optical path prior to being mixed and directed onto the phase reference transimpedance amplifier. Not shown in this figure, the phase interpolator measures changes in phase between the phase reference signal and the measurement signal and calculates the displacement of the target mirror. In our ideal DMI system, a change in phase between these two signals occurs as the interferometer's measurement arm path length only changes with the moving target.

Real

vertical DMI target retro-reflector sandwiched between two moving DMIs
(this assembly is attached to the stage whose displacement is to be measured)

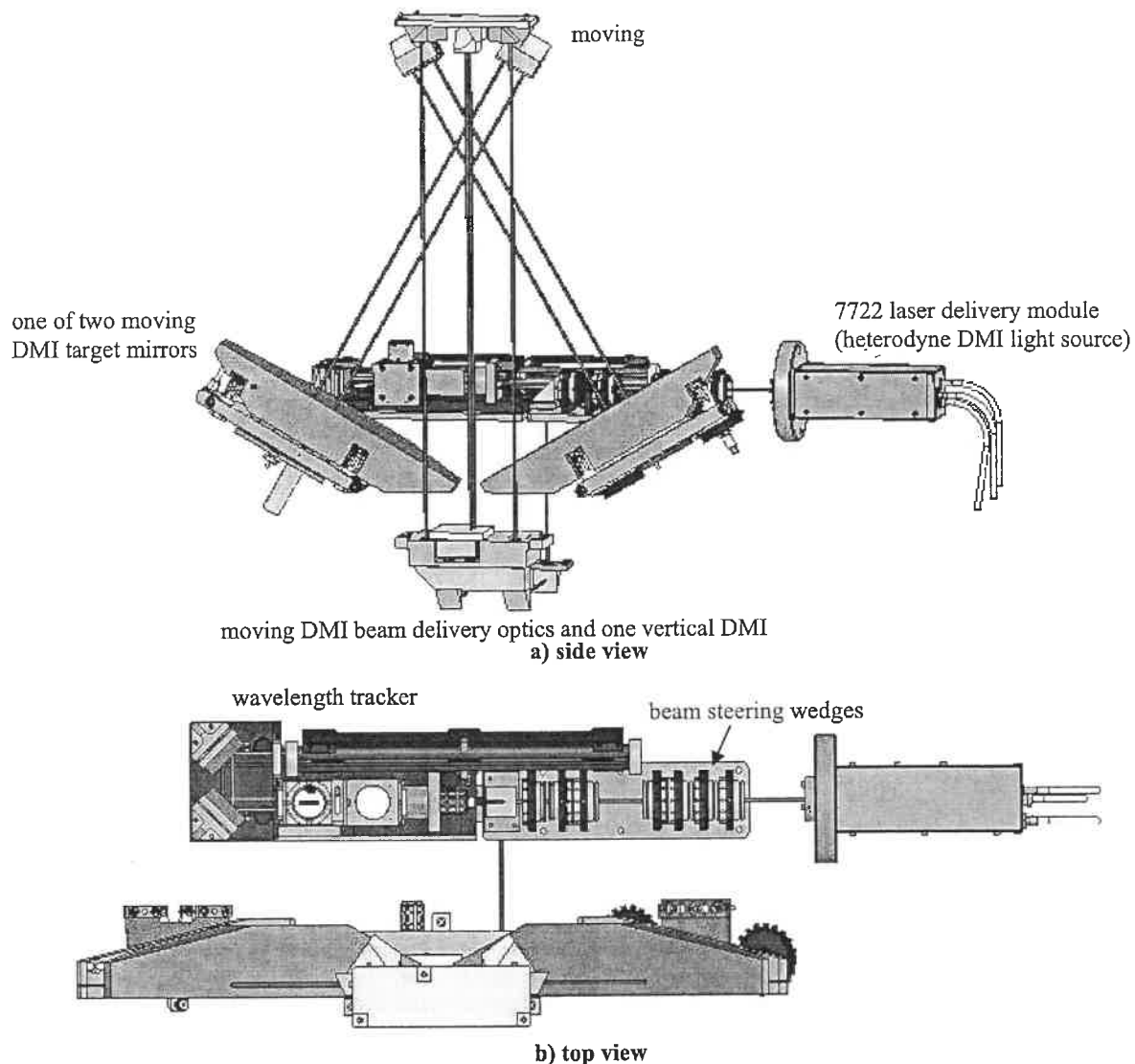


Figure 2 – 4-axis DMI system^{3,4,5}

In reality there are many sources of changing phase of the measurement signal other than target mirror displacement⁶. These other sources contribute to measurement uncertainty and include wavelength instability, beam purity, measurement arm refractive index instability, reference arm refractive index instability, interferometer beam mixing/stray light, wavefront deformation/beam shear, and phase interpolator noise.

Figure 2 shows one of the subsystems – in this case a 4-axis DMI system. The heterodyne DMI light source is a Zygo 7722 laser delivery module⁷. This fiber-fed laser module outputs more than of 1 mW of optical energy while generating less than 20 mW of heat. The output is exactly as described in the previous section with beam purity of 32 ppm (i.e., 32 parts of beam component number 1 with a polarization state matching that of beam component number 2 per 1000000 parts of beam component number 1) and a wavelength instability of <1 ppb over any 1 hour period of time (<1 part-per-billion of the nominal wavelength). The delivery module's optical output is directed to the 4-axes of DMI (3 DMIs and 1 wavelength tracker) via a set of steering wedges. Immediately following the steering wedges, 25% of the beam is allowed to pass to the wavelength tracker while the remaining 75% of the beam is directed downwards to a set of beam steering optics including two right-angle prisms and two rhomboid prisms. The rhomboids split the beam into three beams of “equal” power. One of the beams is directed into a “vertical” DMI while the other two are directed into two “moving” DMIs. Note that in this DMI system the two moving interferometers are moving with the stage motion while their target mirrors are “stationary”. The vertical DMI has a target retro-reflector which moves with the moving interferometers. All three DMIs see a change in the optical path lengths of their measurement arms with vertical motion of the stage to which they are attached (not shown in Figure 2) while only the two moving DMIs see a change in their measurement arm optical path lengths with lateral motions of the stage. With this subsystem we have redundant measurements of the vertical motions of the stage with three distinct optical paths and redundant measurements of the lateral motions with two distinct optical paths. For the purposes of this presentation/abstract, only the measurement uncertainty of one of the moving DMIs will be considered. Further, the measurement uncertainty will be considered in the coordinate frame of the interferometer, not the machine coordinate frame.

The moving DMI

The moving DMI will be operating in a vacuum (about 2 mbar) with a maximum interferometer vertical displacement of 500 mm. The measurement uncertainty is tabulated in Table 1. For the sake of brevity, only measurement uncertainties stemming from interferometer thermal sensitivity and beam sheering of non-planar wavefronts will be considered with some detail here.

uncertainty source	contribution to DMI measurement uncertainty
laser head: beam component purity	0.05 nm
wavelength instability	0.20 nm
moving DMI: thermal sensitivity	0.05 nm
beam mixing/stray light	0.10 nm
beam shear	0.60 nm
phase interpolation electronics	0.15 nm
vacuum index uncertainty	0.32 nm
Total (RSS):	0.73 nm

Table 1- Moving DMI measurement uncertainty (1 σ)

Figure 3 shows the moving interferometer. The reference arm is shown in Figure 3.b while the measurement arm is shown in 3.c. As this interferometer expands and contracts with temperature, it will grow and shrink into and out of the measurement arm some distance proportional to a length and change in temperature. The length used for this calculation is shown in Figure 3.d and is referred to as $l_{\text{sensitive}}$. What we need to do here is to calculate the temperature sensitivity of the optical path length difference between the reference and measurement arms. The optical path length difference is given by

$$\text{OPD} = n_{\text{ref}} \cdot l_{\text{ref}} - n_{\text{meas}} \cdot l_{\text{meas}}$$

where OPD is the optical path length difference between the reference and measurement arms, n_{ref} is the reference arm's index, l_{ref} is the reference arm physical length, n_{meas} is the measurement arm's index, and l_{meas} is the measurement arm physical length. Since the measurement arm consists of both glass and vacuum, it is useful to re-write the equation for optical path length difference as

$$\text{OPD} = n_{\text{SiO}_2} \cdot l_{\text{ref}} - n_{\text{SiO}_2} \cdot l_{\text{meas}_g} - n_v \cdot l_{\text{meas}_v}$$

where n_{SiO_2} is the index of the fused silica, n_v is the index of the vacuum, l_{meas_g} is the measurement arm's physical length in glass, and l_{meas_v} is the measurement arm's physical length in vacuum.

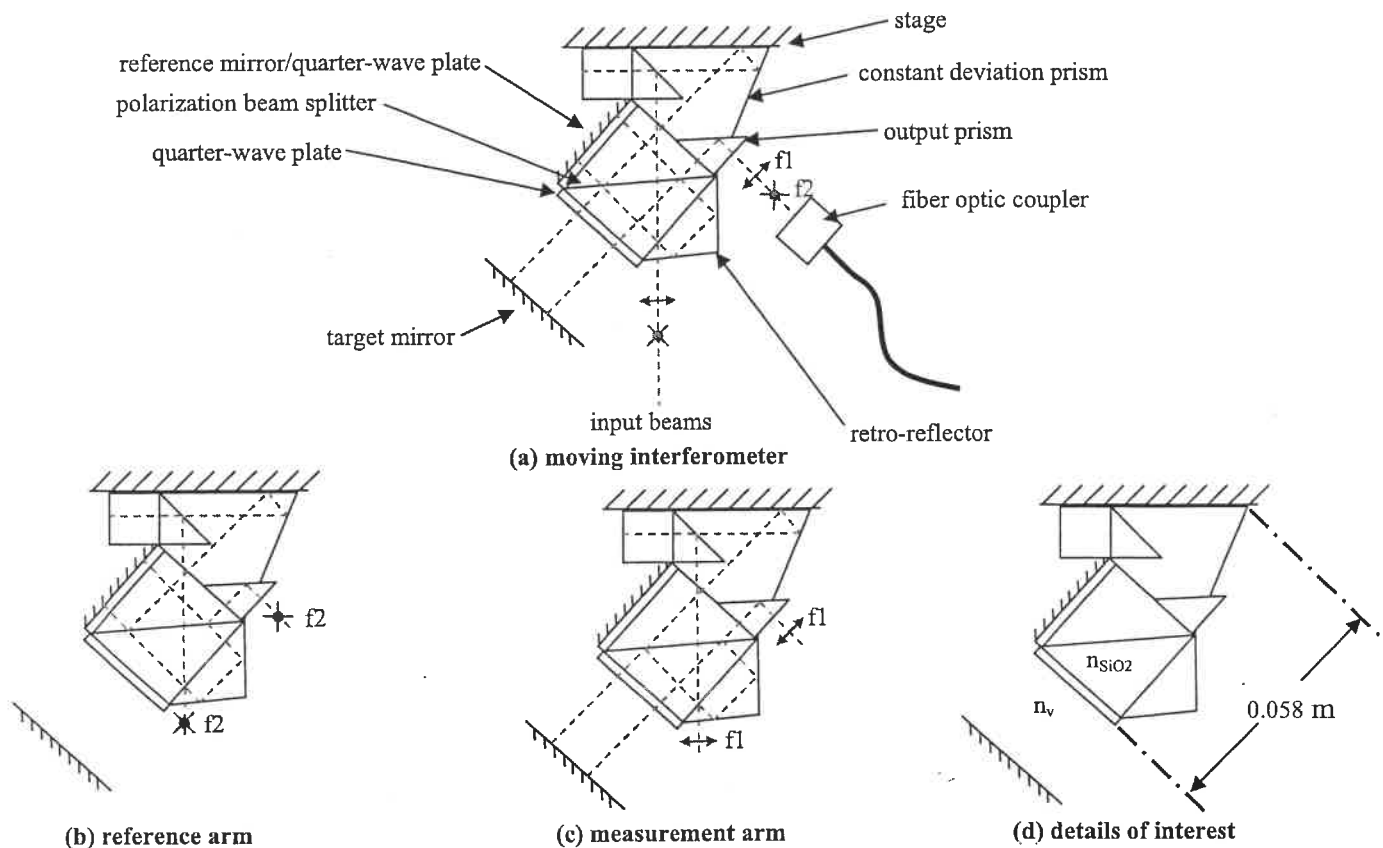


Figure 3 – Moving interferometer schematic diagrams

To find the temperature sensitivity of the interferometer we must solve

$$\frac{d(OPD)}{dT} = \frac{d(n_{SiO_2} \cdot l_{ref} - n_{SiO_2} \cdot l_{meas_g} - n_v \cdot l_{meas_v})}{dT}$$

$$= n_{SiO_2} (l_{ref} - l_{meas_g}) \alpha_{SiO_2} + (l_{ref} - l_{meas_g}) \frac{d(n_{SiO_2})}{dT} - n_v (4 \cdot l_{sensitive}) \alpha_{SiO_2} - l_{meas_v} \frac{d(n_v)}{dT}.$$

We ignore the last term in the equation above since measurement uncertainty due to index variations will be considered/budgeted in the analysis of the wavelength tracker. The input parameters used to solve the above equation are given in Table 2.

model input	description	expectation	standard uncertainty
n_{SiO_2}	index of the fused silica	1.45698	0.000012
$d(n_{SiO_2})/dT$	thermal coefficient for the index of fused silica	8.3 ppm °C ⁻¹	0.3 ppm °C ⁻¹
α_{SiO_2}	thermal expansion coefficient for fused silica	0.55 ppm °C ⁻¹	0.02 ppm °C ⁻¹
n_v	index in the vacuum	1.000001327	7.78 x 10 ⁻⁹
l_{ref}	reference arm physical path length (m)	0.270 m	0.0007 m
l_{meas_g}	measurement arm physical glass path length (m)	0.270 m	0.0007 m
$l_{sensitive}$	sensitive length to thermal growth of interferometer affecting l_{meas_v} (m)	0.058 m	0.0007 m
T	temperature (°C)	23 °C	0.0014 °C
P	pressure (Torr)	3.75 Torr	0.022 Torr
F	partial pressure of the water vapor in the vacuum (Torr)	0 Torr	0 Torr

Table 2 – Model input parameter expected values and standard uncertainties

Plugging in the expected values given in the table above yields an expected interferometer temperature sensitivity of

$$X_{\frac{d(OPD)}{dT}} = 127.6 \text{ nm } ^\circ\text{C}^{-1}.$$

The corresponding standard uncertainty is

$$U_{\frac{d(OPD)}{dT}} = \left[\left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial n_{SiO2}} \right)^2 U_{n_{SiO2}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial n_v} \right)^2 U_{n_v}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial l_{ref}} \right)^2 U_{l_{ref}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial l_{meas_g}} \right)^2 U_{l_{meas_g}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial l_{sensitive}} \right)^2 U_{l_{sensitive}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial \alpha_{SiO2}} \right)^2 U_{\alpha_{SiO2}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial(\frac{dn_{SiO2}}{dT})} \right)^2 U_{\frac{dn_{SiO2}}{dT}}^2 \right]^{\frac{1}{2}}$$

$$= 11.5 \text{ nm } ^\circ\text{C}^{-1}.$$

The uncertainties in glass path length, vacuum path length, and the sensitive ($l_{sensitive}$) length of the interferometer dominate the standard uncertainty. The worst case optical path difference temperature sensitivity is $139.1 \text{ nm } ^\circ\text{C}^{-1}$. Note that this sensitivity has been calculated for the total optical path difference between the reference and measurement arms of the interferometer. Since there are 2 passes to the target mirror in the measurement arm, this number must be divided by 4 to get the measurement sensitivity. Next we multiply the measurement sensitivity by our temperature uncertainty to yield the resulting, conservative, moving DMI measurement uncertainty (1σ) of 0.05 nm .

Beam shear of distorted wavefronts

Some shearing of the measurement arm beam with respect to the reference arm beam within the moving interferometer will occur with changes in stage orientation. Based on the specifics of the application/design, it has been determined that this beam shear will not exceed 0.2 mm . If the measurement and reference arm beam wavefronts were planar perpendicular to the direction of their propagation, then there would be no change in the phase of the heterodyne signal as the measurement arm beam wavefront shears across the reference arm beam wavefront. However, Gaussian beam wavefronts are more spherical than planar. The wavefront departure from planar further increases as the wavefront passes through components in its optical path with inhomogeneous optical properties resulting in parts of the wavefront seeing differing indices than other parts of the wavefront thereby resulting in phase delays across the wavefront, changing its shape. As the non-planar wavefronts are sheared, there is a resulting change in phase of the heterodyne signal which is generated by mixing the electric fields on the face of a fast photodiode and whose phase is the summation of all the little phase differences between the two interfering wavefronts. For modeling purposes, we assume that all beams are Gaussian with complex amplitudes and are well represented by the equation (suppressing the time-dependence of phase)

$$E = \frac{E_0 iz_0}{z + iz_0} \exp \left[-i \frac{n2\pi}{\lambda} \left(\frac{x^2 + y^2}{2(z + iz_0)} + z \right) \right]$$

where E_0 is the electric field amplitude, Z_0 is the Raleigh distance, n is the index, $2\pi\lambda^{-1}$ is the wave number, and x , y , and z are the Cartesian coordinates. Note that the light is propagating in the z -direction. Since our beams are round in cross-section, I note that

$$x = r \cdot \cos(\theta), \quad y = r \cdot \sin(\theta), \quad \text{and} \quad r^2 = x^2 + y^2$$

where r is beam radius and θ is the θ_z . Ultimately we will multiply the sum of the measurement arm beam's electric field and the reference arm beam's electric field with its complex conjugate to get a measurement beam signal. This intensity distribution will be integrated over the active area of the detector to determine the net phase of the signal. Said another way,

$$E_{ref} = \frac{E_0 iz_0}{z_{ref} + iz_0} \exp \left[-i \frac{n_{ref} 2\pi}{\lambda_{ref}} \left(\frac{r_{ref}^2}{2(z_{ref} + iz_0)} + z_{ref} \right) \right],$$

$$E_{meas} = \frac{E_0 iz_0}{z_{meas} + iz_0} \exp \left[-i \frac{n_{meas} 2\pi}{\lambda_{meas}} \left(\frac{r_{meas}^2}{2(z_{meas} + iz_0)} + z_{meas} \right) \right], \text{ and}$$

$$I(r, \theta) = \int_0^r \int_0^{2\pi} (E_{\text{meas}} + E_{\text{ref}}) (E_{\text{meas}} + E_{\text{ref}})^* r dr d\theta.$$

We now use a phase-shifting algorithm to extract phase information from the measurement beam signal, which is proportional to the intensity described by the equation above. In this way we can predict changes in phase that result from beam shearing of the non-planar wavefronts. The magnitude of the DMI measurement error that results from beam shear is a strong function of the shapes of the wavefronts that are being sheared. One can modify the electric field equation to look like

$$E = \frac{E_0}{z + iz_0} \exp \left[-i \frac{n2\pi}{\lambda} \left(\frac{\text{shape}(x, y)}{2(z + iz_0)} + z \right) \right]$$

where $\text{shape}(x, y)$ is a function describing a shape. The really ambitious person could then investigate sensitivities to wavefront shape. It turns out that our wavefronts are predominantly spherical (i.e., $\text{shape}(x, y) = x^2 + y^2$).

moving DMI component	number of passes	specification	expected value	uncertainty
CDP base prism	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
CDP launch	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
PBS cube	4	0.02 λ PV	0.015 λ PV	0.003 λ PV
retro-reflector	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
measurement window/reference mirror	4	0.02 λ PV	0.015 λ PV	0.003 λ PV
output prism	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
steering wedges	6	0.02 λ PV	0.015 λ PV	0.003 λ PV
right angle prism	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
Rhomboid	2	0.02 λ PV	0.015 λ PV	0.003 λ PV
fold mirror	2	0.125 λ PV	0.094 λ PV	0.018 λ PV
delivery module	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
total			0.518 λ PV	0.029 λ PV

Table 3 – Moving DMI component level wavefront distortion specifications

The moving DMI component-level wavefront distortion specifications are given in Table 3. Note that the expected values for the component-level wavefront distortion specifications are assumed correlated as all of the optical components are subjected to similar manufacturing processes. Using the values in Table 3, the measurement error resulting from a 0.2 mm beam shear of the measurement arm beam with respect to the reference arm beam is 0.60 ± 0.04 nm.

Conclusions

Uncertainty analysis plays a vital role in capturing/bounding sources of uncertainty and provides the means by which a successful design can be realized. In this extended abstract, we provide an overview of our approach to reach nanometer level measurement uncertainty with heterodyne displacement measuring interferometry.

References

- 1) Zygo Corporation, "Displacement Measuring Interferometry Primer," www.zygo.com, Jan 1999.
- 2) A. Slocum, "Precision Machine Design," Prentice Hall, Englewoods Cliffs, NJ, 1992.
- 3) Henry Allen Hill, "Apparatus for generating linearly orthogonally polarized light beams," U.S. Patent No. 6157660, Dec. 5, 2000.
- 4) Van Doren, Kuechel, and Evans, "Metrology system for precision 3D motion," U.S. Patent App. 20030223078, Dec. 4, 2003.
- 5) Evans and Zaroni, "Apparatus and methods for high accuracy metrology and positioning of a body," U.S. Patent App. 20030112445, June 19, 2003.
- 6) A.E. Rosenbluth and N. Bobroff, "Optical sources of nonlinearity in heterodyne interferometers," Precision Engineering, 12, 7-11, 1990.
- 7) Zygo Corporation, "ZMI 7722 Laser Source," Operator's Manual OMP-0475A, www.zygo.com, Jan 1999.

Acknowledgements

This work is based on analytical techniques developed at Zygo Corporation over an extended period of time. Some of the key people contributing to these techniques are Dr. Henry Hill, Dr. Peter de Groot, Lars Selberg, Michael Schroeder, and Mike Metz.

Bayesian Methods for Statistical Inference on the Common Mean from Multiple Data Sources

Z.-Q. John Lu Will Guthrie

National Institute of Standards and Technology
Gaithersburg, MD

Abstract

Combining studies or data from different sources for estimation of some related or common quantities has arisen in a number of applied problems. While there are myriad results on the development of various estimators of the common mean using classical/frequentist statistical approaches, there is lack of unifying rules and understanding of the statistical inference procedures such as confidence intervals or hypothesis testing. Indeed, if the variances are different and unknown, the inferential problem is notoriously difficult, and is related to the well-known Behrens-Fisher problem. This paper will show how Bayesian statistics provides a unifying tool for uncertainty analysis in the most general multi-laboratory and multiple methods problems. Software implementations will be discussed for facilitating popularization of Bayesian methods among practical users.

1. Introduction

The problem of combining results from independent studies for the purpose of estimating or testing a common objective has been important in a number of areas, including biomedical studies and metrology. Naturally, there is a large literature in statistics and metrology that deal with this applied problem. Most of the papers in the statistical literature deal with the problem of deriving point estimators of the common mean.

Consider combining two unbiased estimators x_1 and x_2 that have mean μ and variances σ_1^2, σ_2^2 . If the ratio of the variances is known, then one can use the weighted-mean estimator

$$(1) \quad \bar{x} = (\sigma_2^2 x_1 + \sigma_1^2 x_2) / (\sigma_1^2 + \sigma_2^2) = (x_1 / \sigma_1^2 + x_2 / \sigma_2^2) / (1/\sigma_1^2 + 1/\sigma_2^2)$$

with the associated variance

$$\text{var}(\bar{x}) = \sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) = (1/\sigma_1^2 + 1/\sigma_2^2)^{-1}.$$

Note that both of above quantities are known if the ratio of the variances is known since the only unknown variance cancels out. If the two variances are equal, then $\bar{x}$ reduces to the simple-mean estimator: $\bar{x} = (x_1 + x_2)/2$. Indeed, this estimator can perform quite well even in cases where the individual variances are not equal, as its variance, $(\sigma_1^2 + \sigma_2^2)/4$, is smaller than the variance of either individual estimator when $1/3 \leq \sigma_1^2 / \sigma_2^2 \leq 3$. In other words, it is better to combine as long as the variances do not differ too much.

In practice, of course, the variances are almost always unknown. It is precisely because the two variances need to be replaced by some data-based estimators that the problem becomes complicated. For example, Graybill and Deal (1959) show that when the variances are unknown, there does not exist any simple linear estimator of the form: $cx_1 + (1-c)x_2$ for $0 \leq c \leq 1$ which uniformly improves on the individual estimators over all possible values of σ_1^2, σ_2^2 . Thus, the weights in a linearly combined estimator have to be adaptive to the data in order to dominate over the individual estimators. It is natural to replace the unknown variances by their unbiased estimators in (1). For example, if $m_1 s_1^2 / \sigma_1^2$ is distributed as $\chi_{m_1}^2$ and $m_2 s_2^2 / \sigma_2^2$ is distributed as $\chi_{m_2}^2$, then by replacing σ_1^2, σ_2^2 by s_1^2, s_2^2 in (1), Graybill and Deal (1959) showed that the resulting combined estimator is an unbiased estimator and has smaller variance than either x_1 or x_2 when the degree of freedoms (DOFs) satisfy $m_1, m_2 \geq 9$.

In addition, in many real problems, there may be biases associated with each estimator x_1, x_2 used to estimate the common mean μ . In metrology, it is common practice to introduce a bias term, often called a type B uncertainty, in addition to the data-based variance, or type A uncertainty. For example, Eberhardt et al (1989) assumed known

bounds for the biases, and Levenson et al (2000) proposed a data-based estimation method on the bounds for the biases. The effect of introducing a bias term in the individual means is that the resulting estimator is a “shrinkage” weighted mean of the individual estimators and the related variance is inflated in order to incorporate the bias, see Section 2.

The two methods/labs problem can also be generalized to multiple methods or labs. If the number of labs or methods bigger than 5 or 10, the model presented by Rukhin and Vangel (1998), which is similar to the one-way random effects ANOVA model, but with heteroscedastic variances, is a natural framework to develop various estimators of the common mean. Rukhin and Vangel proposed a maximum likelihood estimation method, however, which is difficult to compute. They compared their estimates to various simpler approximate estimates such as the Mandel-Paul algorithm. Note that because the within-lab samples (or DOF associated with each individual mean estimator x_i) may be small in many inter-lab studies, and there will usually be many labs involved (more than 10), the shrinkage or smoothing effect on the common mean inference is more pronounced. Although the Rukhin-Vangel approach and similar approaches using random effects ANOVA as in W.G. Cochran’s approach work well in many cases, they do not address the problem of combining a small number of methods or labs.

In some senses, obtaining a point estimate of the common mean may be a relatively simple problem, since the weighted, shrinkage estimator is often close to the simple mean estimator in many cases. It is the inferential uncertainty, such as confidence intervals or hypothesis testing that appears to be the most complicated issue using the classical/frequentist approach. The reason is that, the distribution of any combined estimator of μ will involve unknown variances as nuisance parameters, so the standard methods have serious limitations in finding exact confidence intervals. For example, Jordan and Krishnamoorthy (1996) proposed some new methods for deriving exact confidence intervals and reviewed some other methods. In general, the classical frequentist statistical approach to the confidence interval is very complicated and often somewhat ad hoc. The related hypothesis testing is no less difficult.

Indeed, when the variances are unknown and unequal, the inferential problem of the common mean is related to the famous Behrens-Fisher problem. In the two-sample case, we have independent vectors $X_1 = (x_{11}, x_{12}, \dots, x_{1m})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$ such that $x_{1i} \sim N(\mu_1, \delta_1^2)$ and $x_{2i} \sim N(\mu_2, \delta_2^2)$ and the sufficient statistics for this problem is: the sample means for each vector, denoted by $\bar{x}_1, \bar{x}_2$ in accordance with previous notations, with variances $\sigma_1^2 = \delta_1^2 / m$, and $\sigma_2^2 = \delta_2^2 / n$, and sample variances for each data vector, s_1^2, s_2^2 . The Behrens-Fisher problem is the inference such as hypothesis testing on the mean difference $\Delta = \mu_1 - \mu_2$. Lee (1989), Chapter 5 gave a clear introduction to this area.

Unlike the classical frequentist statistical approaches, in which the three related problems of point estimation, confidence interval, and hypothesis testing are treated separately, the Bayesian statistical approach offers a more unified approach to the inference problems of the common mean from multiple data sets. Under the Bayesian framework, all inference issues, whether point summaries, credible or HPD intervals, or hypothesis testing, will be embodied in the posterior distribution. The difficulty of accounting for uncertainty due to estimation of nuisance parameters can be more easily dealt with through marginalization, or integration of the full posterior distribution. The Bayesian approach is well suited to problems where there are unknown and unequal variances, and where there are biases in the means that need to be estimated from data. When extra information, such as historical data or expert opinion on the method biases, it can be easily incorporated in the estimate of the common mean using Bayesian methods.

2. Bayesian Formulation of the general p Multi-lab/method Problem

Suppose that the parameters of interest from each lab/method are $\mu_1, \dots, \mu_p$. The data from p labs or methods can be summarized through data reduction using, e.g. standard statistical sufficiency theory: the statistics for the quantities of interest $\mu_1, \dots, \mu_p$ are respectively, $x_1, \dots, x_p$, with associated variances $\sigma_1^2, \dots, \sigma_p^2$. More precisely, we make the assumption that

$$(2) \quad x_1 \sim N(\mu_1, \sigma_1^2), \dots, x_p \sim N(\mu_p, \sigma_p^2), \text{ mutually independent,}$$

And that there exist Chi-squared random variables $s_1^2, \dots, s_p^2$ such that

$$(3) \quad m_1 s_1^2 / \sigma_1^2 \sim \chi_{m_1}^2, \dots, m_p s_p^2 / \sigma_p^2 \sim \chi_{m_p}^2$$

where $m_1, \dots, m_p$ are the degrees of freedom (DOFs) of the individual variance estimates from each lab. We also assume that $s_1^2, \dots, s_p^2$ are independent of $x_1, \dots, x_p$.

The parameters of interest from each lab/method $\mu_1, \dots, \mu_p$ are related to the *common* parameter of interest μ through the equations

$$(4) \quad \mu_1 = \mu + b_1, \dots, \mu_p = \mu + b_p.$$

Here the b_i 's are lab biases, which are assumed to have normal distribution with zero mean and variance s^2 .

The question now is to obtain estimates for unknown parameters of interest, $\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2$. By the Bayesian theorem, that the posterior is proportional to the product of likelihood function and the prior distribution on all unknown parameters. By the fact that the lab mean parameters may be integrated out from (2) and (4), the joint posterior distribution of the remaining unknown parameters $[\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2 | x_1, \dots, x_p; s_1^2, \dots, s_p^2]$ is proportional to:

$$(5) \quad \begin{aligned} & \propto \prod_{i=1}^p [x_i | \mu, \sigma^2, \sigma_i^2] \cdot \prod_{i=1}^p [s_i^2 | \sigma_i^2] \cdot [\mu] [\sigma_1^2, \dots, \sigma_p^2; \sigma^2] \\ & = \prod_{i=1}^p \exp\left[-\frac{(x_i - \mu)^2}{2(\sigma^2 + \sigma_i^2)}\right] \cdot \prod_{i=1}^p (\sigma^2 + \sigma_i^2)^{-1/2} \cdot \\ & \quad \prod_{i=1}^p (\sigma_i^2)^{-\frac{m_i}{2}} \exp\left[-\frac{m_i s_i^2}{2\sigma_i^2}\right] \cdot [\mu] \prod_{i=1}^p [\sigma_i^2] [\sigma^2 | \sigma_i^2] \end{aligned}$$

where the last term $[\mu] \prod_{i=1}^p [\sigma_i^2] [\sigma^2 | \sigma_i^2]$ is from the prior distribution assumption.

The joint posterior distribution (5) embodies all information that is needed to make inference on the unknown parameters $\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2$. For example, the Bayesian estimators of $\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2$ based on the mode of the posterior distribution are clearly related to the maximum likelihood estimators of Rukhin and Vangel (1998). But the Bayesian approach also easily handles situations when restrictions on the variance components, such as $\sigma^2 \geq 0, \sigma_1^2 > 0, \dots, \sigma_p^2 > 0$, may be in effect. More importantly, Bayesian approach offers a clear way of incorporating the uncertainty in the unknown nuisance parameters $\sigma^2, \sigma_1^2, \dots, \sigma_p^2$ in the assessment of uncertainty in the common mean inference.

Computation based on (5) is significantly simplified under balanced design assumptions, i.e. when $\sigma_1^2 = \dots = \sigma_p^2$, a natural consequence from exchangeability considerations among labs/methods. Then (5) specializes to

$$(6) \quad [\mu, \sigma^2, \sigma^2 | \text{data}] \propto (\sigma_1^2)^{-\frac{1}{2}n} (\sigma^2 + \sigma_1^2)^{-\frac{1}{2}p} \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^p m_i s_i^2}{\sigma_1^2} + \frac{\sum_{i=1}^p (x_i - \bar{x})^2}{\sigma^2 + \sigma_1^2} + \frac{p(\bar{x} - \mu)^2}{\sigma^2 + \sigma_1^2}\right]\right\} \\ \cdot [\mu] [\sigma^2 | \sigma_1^2] [\sigma_1^2], \quad -\infty \leq \mu \leq \infty, \quad \sigma_1^2 > 0, \quad \sigma^2 > 0,$$

where $v_1 = \sum_{i=1}^p m_i$, $\bar{x} = \frac{1}{p} \sum_{i=1}^p x_i$. To obtain the posterior distribution of the variance components (σ^2, σ_1^2) , (6) is integrated over μ yielding

$$(7) \quad [\sigma_1^2, \sigma^2 | \text{data}] \propto (\sigma_1^2)^{-\frac{1}{2}v_1} (\sigma^2 + \sigma_1^2)^{-\frac{1}{2}(p-1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^p m_i s_i^2}{\sigma_1^2} + \frac{(p-1)s_b^2}{\sigma^2 + \sigma_1^2} \right] \right\} [\sigma^2 | \sigma_1^2] [\sigma_1^2]$$

$$\sigma_1^2 > 0, \sigma^2 > 0,$$

Where $s_b^2 = \frac{1}{p-1} \sum_{i=1}^p (x_i - \bar{x})^2$ is an estimate of the between lab/method variability. Estimation of σ^2 and σ_1^2 should be done jointly, because of the constraints $\sigma_1^2 > 0$, $\sigma^2 > 0$ and the imprecise nature in the inter-lab variance statistics s_b^2 . Box and Tiao (1973, pp. 252-264) amply demonstrated this point. Overall, computation of (7) is fairly straightforward to implement, either numerically, or by Monte Carlo using χ^{-2} distribution representation as used by Box and Tiao. But, because of the heavy-tailed nature of the marginal posterior distribution of σ^2 when p is small, say ≤ 10 , then it is preferable to obtain the posterior distribution of the common mean parameter μ by direct integration of (6) over $\sigma_1^2 > 0$, $\sigma^2 > 0$. Alternatively, one can write the posterior of μ as an expectation of familiar densities over (7):

$$(8) \quad [\mu | \text{data}] = E_{[\sigma^2, \sigma_1^2 | \text{data}]} \Phi(\bar{x}, \delta^2) [\mu],$$

where Φ is the normal density function, and $\delta^2 = \frac{1}{p} (\sigma^2 + \sigma_1^2)$.

3. Prior Choices

For simplicity, we assume independent priors on all unknown parameters i.e. $[\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2] = [\mu][\sigma^2] \prod_{i=1}^p [\sigma_i^2]$. When the number of observations for each method/lab is significant, estimation of individual variance components is relatively easy and is not strongly influenced by prior choices. Furthermore, estimation of μ itself is relatively stable, so we will focus on the prior choices for σ^2 mainly.

Noninformative choices are popular since they require no prior knowledge. Among the potential choices, it turns out that one has to be extremely careful about a common prior choice for variance $\pi_1(\sigma^2) = 1/\sigma^2$ or the related vague prior using inverse Gamma $\pi_1(\sigma^2) = (\sigma^2)^{-(\varepsilon+1)} \exp(-\lambda/\sigma^2)$ when ε, λ are small. Such vague prior choice for leads to problematic issues for the posterior distribution. Indeed, such a prior choice may lead to an improper or close to improper posterior distribution. If the Laplace noninformative prior $\pi_2(\sigma^2) \equiv 1$ is used, then the posterior distribution is not proper for $p = 2$, and is only a proper density function when $p \geq 3$.

As a general recommendation, we suggest the noninformative prior:

$$(9) \quad \pi_3(\sigma^2) \propto 1/(c + \sigma^2),$$

where c is some constant to be specified, as also recommended in Box and Tiao (1973, p.372). When all labs have comparable variances: $\sigma_1^2 = \dots = \sigma_p^2 = \bar{\sigma}_a^2$, a natural choice is to take $c = \bar{\sigma}_a^2$. It can be shown that the posterior of the inter-lab variance is a proper density function, thanks in part to the partially informative prior $\pi_3(\sigma^2) \propto 1/(\bar{\sigma}_a^2 + \sigma^2)$. However, the posterior mean (and variance) of σ^2 tends to infinity when $p \leq 2$ (or $p \leq 4$) because the tail of $p_3(\sigma^2)$ decays as $(\sigma^2)^{-(p/2+1)}$ when $\sigma^2 \rightarrow \infty$. The dependence of this prior choice on the within-lab variance σ_a^2 may be reasonable, considering the fact that estimation of σ^2 , and σ_a^2 is indeed, closely correlated. Box and Tiao (1973, Chapter 3) gave a very detailed discussion of

Bayesian inference using $\pi_3(\sigma^2) \propto 1/(c + \sigma^2)$, and showed that how the Bayesian approach easily overcame some of the complications of negative estimates of σ^2 which have plagued the classical “point estimation” approaches for a long time. Though recent developments of Markov chain Monte Carlo (Robert and Casella 1999) have taken away some of the needs of clever integral manipulations in the traditional Bayesian calculations as discussed in Box and Tiao (1973), we think for the heavy-tailed posterior calculations, the direct numerical integration approach is still much needed, for calculating the posterior density function and HPD credible intervals.

Apart from the noninformative choices, one can certainly consider informative prior choices in order to model expert opinions or historical data. For example, Levenson et al (2000) has used an uniform prior on the between-lab bias. The conjugate prior choice using inverse Gamma is flexible approach for modeling the variance of between-lab bias. That is,

$$(10) \quad \sigma^2 \sim \tau^2 \chi_k^{-2}, \text{ or } IG(\tfrac{1}{2}k, \tau^2).$$

Here we use IG to denote the inverse Gamma distribution and $IG(\tfrac{1}{2}k, \tau^2)$ of degree of freedom $\tfrac{1}{2}k$ and scale τ^2 has the density function of the form:

$$\pi_4(\sigma^2) = \frac{(\tau^2)^{k/2}}{2^{k/2} \Gamma(k/2)} (\sigma^2)^{-(\frac{1}{2}k+1)} \exp\left(-\frac{\tau^2}{2\sigma^2}\right), \sigma^2 > 0.$$

The general setup is easily amenable to a Markov Chain Monte Carlo implementation through (5). To obtain the marginal distribution of μ, σ^2 , one can simply draw random samples of $\sigma_1^2, \dots, \sigma_p^2$ from inverse chi-squares, then use Gibbs sampling to generate (dependent) samples through μ , and σ^2 alternatively.

4. Examples

The proposed Bayesian methods have been applied to some typical common mean inference problems in the literature, including the Eberhardt et al (1989) example, also studied by Jordan and Krishnamoorthy (1996). Another example is a non-Gaussian example, in which one needs to combine binomial trials using a Bayesian Beta-binomial model in order to account for the between-sample variability.

References

- [1] Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Inference*. Addison-Wesley, Reading, Massachusetts.
- [2] Eberhardt, K.R., Reeve, C.P. and Spiegelman, C.H. (1989). A Minimax Approach to Combining Means, With Practical Examples. *Chemometrics and Intelligent Laboratory Systems*, **5**, 129-148.
- [3] Graybill, F.A. and Deal, R.B. (1959). Combining Unbiased Estimators. *Biometrics*, **15**, 543-550.
- [4] Jordan, S.M. and K. Krishnamoorthy (1996). Exact Confidence Intervals for the Common Mean of Several Normal Populations. *Biometrics*, **52**, 77-86.
- [5] Lee, P. M. (1989). *Bayesian Statistics: An Introduction*. Oxford University Press, New York.
- [6] Levenson, M.S., D.L. Banks, K.R. Eberhardt, L.M. Gill, W.F. Guthrie, H.K. Liu, M.G. Vangel, J.M. Yen, and N.F. Zhang (2000). An Approach to Combining Results from Multiple Methods Motivated by the ISO GUM. *J. Res. Natl. Ints. Stand. Technol.*, **105**, No.4, 571-579.
- [7] Rukhin, A.L. and M.G. Vangel (1998). Estimation of a Common Mean and Weight Means Statistics. *Journal of American Statist. Association*, **93**, No.441, 303-308.

Uncertainty Analysis of Interlaboratory Studies with Linear Trends

Nien F. Zhang, Hung-kung Liu, Nell Sedransk, and William E. Strawderman
National Institute of Standards and Technology, Gaithersburg, MD

Interlaboratory studies have been used to ensure accurate measurement capability among different laboratories. In this paper, we discuss a special interlaboratory comparison called key comparison. With the recent signing of the Mutual Recognition Arrangement (MRA), national metrology institutes (NMIs) and regional metrology organizations (RMOs) around the world have committed themselves to establishing the equivalence of their measurement standards through key comparisons of national measurement standards. In most key comparisons, the pilot NMI is responsible to organizing the circulation and transport of standards or artifacts and ensuring that the participant NMIs make proper arrangements. See the Appendix F of the "Mutual Recognition of National Measurements Standards and of Measurement Certificates issued by National Metrology Institutes" [1]. Usually each non-pilot NMI makes measurements in one period while the pilot NMI makes measurements in multiple periods based on the circulation scheme. In some key comparisons, the measurements of the transport standards made by the participating NMIs will show a trend or a drift (a linear drift in this paper). If this occurs, a traditional statistical analysis, which assumes no trend exists as in the recent publications, is inappropriate. Within some consultative committees (CCs) of the International Committee for Weights and Measures (CIPM), some uncertainty analyses are performed to accommodate the drift effect. One approach is for each NMI to calculate the difference between the measured value and the predicted value based on the trend, e.g., a linear trend. By combining these differences, e.g., using a weighted mean, the key comparison reference value (KCRV) is then calculated. However, in the process of calculating the uncertainties relating to the KCRV and other quantities, the correlations among the predicted values of the non-pilot NMIs, which are jointly based on the linear regression from the measured values of the pilot NMI, were not accounted for and thus a more rigorous statistical study was needed. In the paper by Zhang, Sedransk, and Jarrett [2], an approach was proposed and applied to the CCEM-K2 Key Comparison of Resistance Standards at 10 M Ω and 1 G Ω (see Dziuba and Jarrett [3]). The proposed method in the above article improved the uncertainty analysis by considering the correlations (which were ignored in many practices) due to the predictions for non-pilot laboratories jointly based on the pilot laboratory measurements. However, due to the definition of KCRV in the existing methods dealing with trend, the results cannot be applied to the trivial case when the trend reduces to zero. In this article, we define the time-dependent KCRV in an appropriate way and propose a new approach that overcomes the shortcomings in [2] and [3]. The calculation of KCRV is consistent with the case in which there is no trend. The pair-wise comparison between any two national measurement standards and its corresponding uncertainty, which is independent of the time, are also given.

We assume that the measurements of any particular laboratory have a linear trend in time and the slopes of the linear trends for all the laboratories are the same while we allow for different intercepts for different laboratories. In this article we consider the case of a

single artifact circulating through all laboratories. Without loss of generality, we assume that the pilot laboratory is the first one among all I laboratories. Denote the time and the result of the k^{th} measurement by the pilot laboratory by t_{1k} , $k=1,\dots,K$, which are taken to have negligible associated uncertainty, and X_{1k} , $k=1,\dots,K$ with $K>2$, respectively. In practice, t_{1k} can be an average value of the time when the measurements were made in the k^{th} period and then X_{1k} is the average value of the corresponding measurements in that period.

We assume that a simple linear regression model holds for the measurements made by the pilot laboratory,

$$X_{1k} = \alpha_1 + \beta t_{1k} + \varepsilon_{1k} \quad (1)$$

for $k=1,\dots,K$. We assume that the random error ε_{1k} has a zero mean and uncertainty of u_1 . When $i \neq 1$, each laboratory takes one measurement at time t_i and the corresponding model is

$$X_i = \alpha_i + \beta t_i + \varepsilon_i, \quad (2)$$

where the random error ε_i has a zero mean and standard uncertainty of u_i for $i=2,\dots,I$. In a key comparison study, NMI's will provide x_i 's, the values of X_i 's, and the corresponding uncertainties including the Type A and Type B evaluations of uncertainty.

For the pilot laboratory, the least squares estimators of the regression parameters are given by

$$\hat{\beta} = \frac{\sum_{k=1}^K (t_{1k} - t_1)(X_{1k} - X_1)}{\sum_{k=1}^K (t_{1k} - t_1)^2} \quad (3)$$

and

$$\hat{\alpha}_1 = X_1 - \hat{\beta} t_1, \quad (4)$$

where X_1 is the average of the measurements made by the pilot laboratory and t_1 is the average of $\{t_{11}, \dots, t_{1K}\}$. For other laboratories, i.e., $i=2,\dots,I$, the corresponding intercepts can be estimated by

$$\hat{\alpha}_i = X_i - \hat{\beta} t_i, \quad (5)$$

where t_i is the time when the i^{th} ($i \neq 1$) laboratory made its measurement.

Key Comparison Reference Value

One approach to calculating the KCRV at any time t (denoted by $KCRV_t$) is to use a weighted mean of $\hat{\alpha}_i + \hat{\beta}t$ over the laboratories $i = 1, \dots, I$,

$$KCRV_t(w) = \sum_{i=1}^I w_i(\hat{\alpha}_i + \hat{\beta}t) = \sum_{i=1}^I w_i X_i - \hat{\beta} \sum_{i=1}^I w_i(t_i - t) \quad (6)$$

with weights $w = (w_1, \dots, w_I)$ satisfying $\sum_{i=1}^I w_i = 1$. $KCRV_t(w)$ is a function of t as well as the weights. The uncertainty of KCRV depends on t as well as the weights w . For any given t , a reasonable criterion to choose optimal weights is the minimum variance or minimum standard uncertainty of $KCRV_t(w)$. For any given set of weights, the minimum value of the standard uncertainty of $KCRV_t(w)$ will occur when $\sum_{i=1}^I w_i(t_i - t)^2$ equals zero. This occurs when t takes the value of $t_w = \sum_{i=1}^I w_i t_i$. In this case, from (6), the KCRV has the same form as in the no trend case and is given by

$$KCRV_{t_w}(w) = \sum_{i=1}^I w_i X_i. \quad (7)$$

The corresponding uncertainty is given by

$$u_{KCRV_{t_w}(w)}^2 = \sum_{i=1}^I w_i^2 u_i^2. \quad (8)$$

As in the no trend case, $u_{KCRV_{t_w}(w)}^2$ is minimized when the weights are given by

$$w_i(1) = \frac{\frac{1}{u_i^2}}{\sum_{k=1}^I \frac{1}{u_k^2}}. \quad (9)$$

The corresponding KCRV is given by

$$KCRV_{t^*}(w(1)) = \sum_{i=1}^I w_i(1) X_i, \quad (10)$$

where

$$t^* = \sum_{i=1}^I w_i(1) t_i, \quad (11)$$

and $w(1) = (w_1(1), \dots, w_I(1))$. From (9) and (10), the uncertainty of this KCRV is the same as in the no-trend case and is given by

$$u_{KCRV_i^*(w(1))}^2 = \frac{1}{\sum_{i=1}^I \frac{1}{u_i^2}} \quad (12)$$

assuming that u_i ($i=1, \dots, I$) in the weights $\{w_i(1)\}$ are the standard deviations for all laboratories. As a practical matter, we recommend the $KCRV_i^*(w(1))$ with t^* and the weights given by (9) and (11) to be the reference value of this key comparison.

Degrees of equivalence of the national measurement standards with respect to the KCRV

The degree of equivalence of the national measurement standard from the i^{th} laboratory with respect to the $KCRV_i$ is defined as the difference:

$$\begin{aligned} D_{iKCRV(w)} &= \hat{\alpha}_i + \hat{\beta}t - KCRV_i(w) \\ &= (1 - w_i)\hat{\alpha}_i + \sum_{j \neq i, j=1}^I w_j \hat{\alpha}_j. \end{aligned} \quad (13)$$

when $w = w(1)$ as given in (9), in the first equality of (13), let $t = t^* = \sum_{i=1}^I w_i(1)t_i$. Then

$$\begin{aligned} D_{iKCRV(w(1))} &= \hat{\alpha}_i + \hat{\beta}t^* - KCRV_i^*(w(1)) \\ &= \hat{\alpha}_i + \hat{\beta}t^* - \sum_{i=1}^I w_i(1)X_i. \end{aligned} \quad (14)$$

The corresponding uncertainty is given by

$$u_{D_{iKCRV(w(1))}}^2 = [1 - 2w_i(1)]u_i^2 + \frac{1}{\sum_{k=1}^I \frac{1}{u_k^2}} + \frac{(t_i - t^*)^2 \sigma_{1,A}^2}{\sum_{k=1}^K (t_{1k} - t_1)^2}, \quad (15)$$

where $\sigma_{1,A}$ is the Type A evaluation of uncertainty of the measurements made by the pilot laboratory. An unbiased estimator of $\sigma_{1,A}^2$ in (15) is given by the “mean-square residual” based on the model in (1), i.e.,

$$\hat{\sigma}_{1,A}^2 = \frac{\sum_{k=1}^K (X_{1k} - \hat{\alpha}_1 - \hat{\beta}t_{1k})^2}{K - 2}. \quad (16)$$

Again we recommend using $D_{iKCRV(w(1))}$ as the degree of equivalence of the national measurement standard from the i^{th} laboratory with respect to the KCRV.

Degrees of equivalence of pairs of national measurement standards

The degree of equivalence between the national measurement standards at time t is defined as

$$D_{i,k} = (\hat{\alpha}_i + \hat{\beta}t) - (\hat{\alpha}_k + \hat{\beta}t) = \hat{\alpha}_i - \hat{\alpha}_k \quad (17)$$

when $i \neq k$. Thus, the quantity is independent of t . The corresponding uncertainty is

$$u_{D_{i,k}}^2 = u_i^2 + u_k^2 + \frac{(t_i - t_k)^2 \sigma_{1,A}^2}{\sum_{k=1}^K (t_{1k} - t_1)^2} \quad (18)$$

To illustrate the approach, we applied it to the key comparison CCEM-K2. In this key comparison, three 10 M Ω wirewound resistors and three 1 G Ω film-type resistors were used as the transport standards. During the comparison, the transport standards were measured at the pilot NMI, NIST. For each period an average value of the dates when the measurements were made is called a mean date of the period. Each non-pilot NMI reported a mean value and a mean date of measurement for each of the six artifacts. An uncertainty budget that includes the Type A and Type B evaluations of uncertainties for each NMI's measurement process was also reported. In this example, we only consider the 10 M Ω case and use the measurement data only for the artifact, S/N HR7551 listed in Table 1. The uncertainties for all NMIs are also listed in Table 1.

For the trend, the assumption is that the resistor drifts in a linear fashion. Any non-linear effects are caused probably by several physical or mechanical changes during the transportation process. A linear regression line was fit to the NIST measurements. The estimates of the intercept for NIST, the slope, and the residual standard deviation are $\hat{\alpha}_1 = 6.03$, $\hat{\beta} = 1.05$, and $\hat{\sigma}_{1,A} = 1.09$, respectively. We used $\hat{\sigma}_{1,A}$, the residual standard deviation from (16) as an estimate of $\sigma_{1,A}$ in (15) and (18). The date for minimum standard deviation of KCRV is $t^* = 1998.23$ given by (11). The KCRV at t^* with weights $w(1)$ given by (9) and (10) is 8.03 and the corresponding uncertainty is 0.28 given by (12). The $D_{iKCRV,t^*(w(1))}$ for all NMIs and their corresponding uncertainties $u_{D_{iKCRV,t^*(w(1))}}$ were calculated.

The degrees of equivalence of pairs of NMIs and their corresponding uncertainties are calculated by (17) and (18), respectively.

In summary, in this paper we propose a new statistical analysis for key comparisons with linear trends. The calculation of KCRV is consistent with the case in which there is no

trend. The corresponding uncertainties for KCRV and degrees of equivalence are also provided.

1. Guidelines for CIPM key comparisons (Appendix F to *Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes*), Technical Report, International Committee for Weights and Measures, 1999.

2. Zhang, N. F., Sedransk, N., and Jarrett, D. G., Statistical Uncertainty Analysis of Key Comparisons CCEM-K2, *IEEE Transaction Instrumentation & Measurement*, 2003, 491-494.

3. Dziuba, R. F. and Jarrett, D. G. Final Report CCEM-K2 Key Comparison of Resistance Standards at 10 M Ω and 1G Ω . Appendix B of *Mutual Recognition Arrangement*, 2002, Bureau International des poids et Mesures (BIPM), Paris France.

Table 1. Information for the Standard S/N HR7551

Lab	Mean date of Measurement	Measurement (x 10 ⁻⁶)	Type A uncertainty (x 10 ⁻⁶)	Type B uncertainty (x 10 ⁻⁶)
NIST	1996.65	4.6	0.2 *	1.51*
NRC	1996.80	5	1.88	2.29
NIST	1996.94	6.7	0.2 *	1.51*
BNM-LCIE	1997.17	6.97	0.50	0.35
NPL	1997.35	7.1	0.52	0.61
PTB	1997.50	7.5	1.0	2.19
NIST	1997.62	8.1	0.2 *	1.51*
CSIRO-NML	1997.82	7.3	0.07	2.56
MSL	1998.03	7.3	0.04	0.59
CSIR-NML	1998.13	-20	50	13.52
NIST	1998.33	8.9	0.2 *	1.51*
SP	1998.49	8.7	0.17	1.79
OFMET	1998.62	8.9	0.39	0.58
IEN	1998.74	9.4	0.79	2.53
NMi-VSL	1998.98	9.1	0.80	3.04
NIST	1999.15	7.8	0.2 *	1.51*
KRIST	1999.39	7.1	0.30	3
NIST	1999.60	8.5	0.2 *	1.51*
NIM	1999.87	10.2	0.10	0.83
VNIIM	2000.03	10	0.25	1.03
NIST	2000.20	10	0.2 *	1.51*

Uncertainty Analysis of Kolsky Bar Strain Gage Output

Richard Rhorer
National Institute of Standards and Technology
Gaithersburg, MD 20899-8223

An apparatus has been built at the National Institute of Standards and Technology (NIST) to measure the dynamic stress-strain relationships for metals at elevated temperatures. The NIST Pulse Heated Kolsky Bar apparatus has the capability to resistively heat samples to over 1000 °C with heating rates of over 10 000 °C/s in addition to performing the traditional Kolsky bar impact testing. The traditional Kolsky bar (or Split Hopkinson Pressure Bar) consists of two long hardened steel bars aligned end-to-end with the sample sandwiched in between the bars. A test is carried out by striking the end of the incident bar with a projectile fired from an air gun that sends an elastic wave down the bar, which then rapidly deforms the sample. As the sample is deformed a transmitted wave is started down the second bar. Standard variable resistance metal foil strain gages are mounted in the center of each of the long bars that measure the amplitude of the strain wave as a function of time. The outputs of the strain gages are recorded on a standard digital oscilloscope data acquisition system. Using the strain gage signal recordings a high-strain rate stress-strain curve for the sample material can then be calculated using a well established method. Our Kolsky bar apparatus can test samples at strain rates from about 500 /s to 10 000 /s (normal compression machines in materials test labs operate in the range of 0.01 /s and under special conditions can go as high as 100 /s.) By measuring the temperature of the sample during the deformation (the deformation takes approximately 150 microseconds) the stress-strain curve can be reported for various temperatures.

The strain gage measuring system is calibrated by a standard method of shorting a large precision resistor across the gage and recording the output. The uncertainty of the simulated strain in this parallel resistor calibration method depends on the uncertainty of the value of the parallel resistor, the gage resistance, and the manufacturer's gage factor. There is an additional contribution to the uncertainty of the measurement results because of the dynamic effects on the gage factor. To understand the accuracy of the strain gage output the different uncertainties need to be combined in an uncertainty analysis. The uncertainty analysis of the strain gage calibration method presented in this paper uses an approach based on NIST Technical Note 1297. Future papers will address the uncertainty of the temperature measuring methods and some of the effects of the assumptions used in the Kolsky bar simple theory for relating strain in the elastic bars to the stress and strain in the sample.

DISPLACEMENT MEASURING INTERFEROMETRY MEASUREMENT UNCERTAINTY

Mike Holmes, Chris Evans
Zygo Corporation, Middlefield, CT 06455

Introduction

Displacement measuring interferometry is a key enabling technology allowing the semiconductor industry to continue to shrink linewidths by a factor of two every 2-3 years and increase integrated circuit performance. Intel, for example, have indicated that they expect to see limited production at the 65 nm node in 2005; line widths down to 32 nm by the end of the decade are predicted in the most recent ITRS roadmap. Typically, overlay budgets are about 1/3 the linewidth and reticle and wafer stage metrology themselves are but one of many factors contributing to overlay budget. Hence it is clear that sub-nm wafer stage position uncertainties will be needed on the stages in the lithography tools used to make the integrated circuits.

Nanometer class position uncertainties are required in other applications, such as a long travel measuring system being developed at Zygo. In this application, the position of a 6 degree of freedom stage with a working volume of 500 x 1 x 1 mm must be known with an uncertainty of 1 nm ($k=1$). Including calibration tools and the wavelength tracker, 14 DMIs are used in four different configurations. This presentation focuses on one of those configurations, and gives a detailed analysis of the uncertainties; the methodology is shown here for two uncertainty contributors.

Ideal

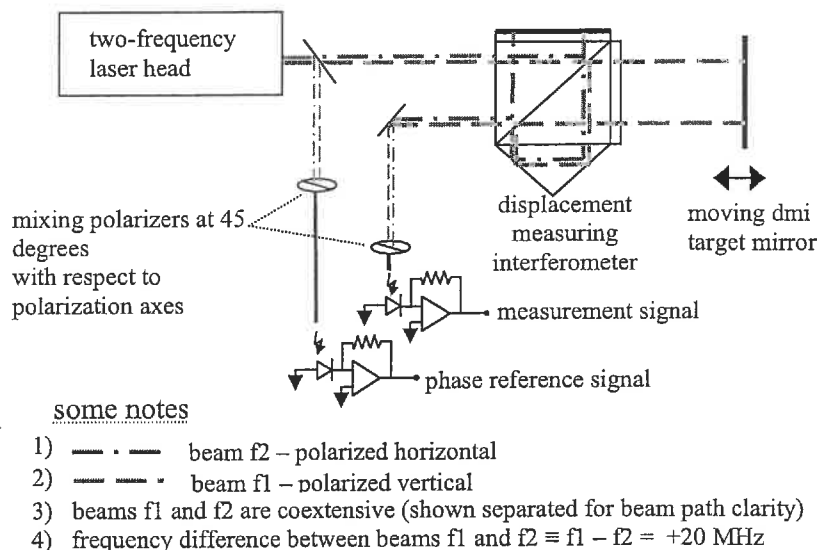


Figure 1 – Ideal heterodyne displacement measuring interferometer system minus phase interpolator

Figure 1 shows an idealized heterodyne Displacement Measuring Interferometer (DMI) system^{1,2} consisting of the laser head, displacement measuring interferometer, moving target mirror, and transimpedance amplifiers. The output from the laser head is two coextensive linearly polarized beams ($\lambda_{1,2} \approx 632.991$ nm) which are separated in frequency by 20 MHz. The polarization states are orthogonal to each other. The two components are separated in the interferometer. One component travels a constant optical path through the reference arm of the interferometer while the other travels through the measurement arm whose optical path length is dependent upon the position of the target mirror. The two components rejoin inside the interferometer and are mixed and directed onto the measurement signal transimpedance amplifier. The electrical output from this amplifier is a $20 \text{ MHz} \pm \text{Doppler shift}$ (a Doppler shift occurs when the target mirror is moving and is proportional to its velocity) sinusoid. This signal's magnitude is

proportional to the heterodyne signal optical power and its phase is proportional to the interferometer's measurement arm optical path length. The electrical phase reference signal is generated from a portion of the interferometer's input beams. Note that this signal has constant phase as the two components are traveling a common constant optical path prior to being mixed and directed onto the phase reference transimpedance amplifier. Not shown in this figure, the phase interpolator measures changes in phase between the phase reference signal and the measurement signal and calculates the displacement of the target mirror. In our ideal DMI system, a change in phase between these two signals occurs as the interferometer's measurement arm path length only changes with the moving target.

Real

vertical DMI target retro-reflector sandwiched between two moving DMIs
(this assembly is attached to the stage whose displacement is to be measured)

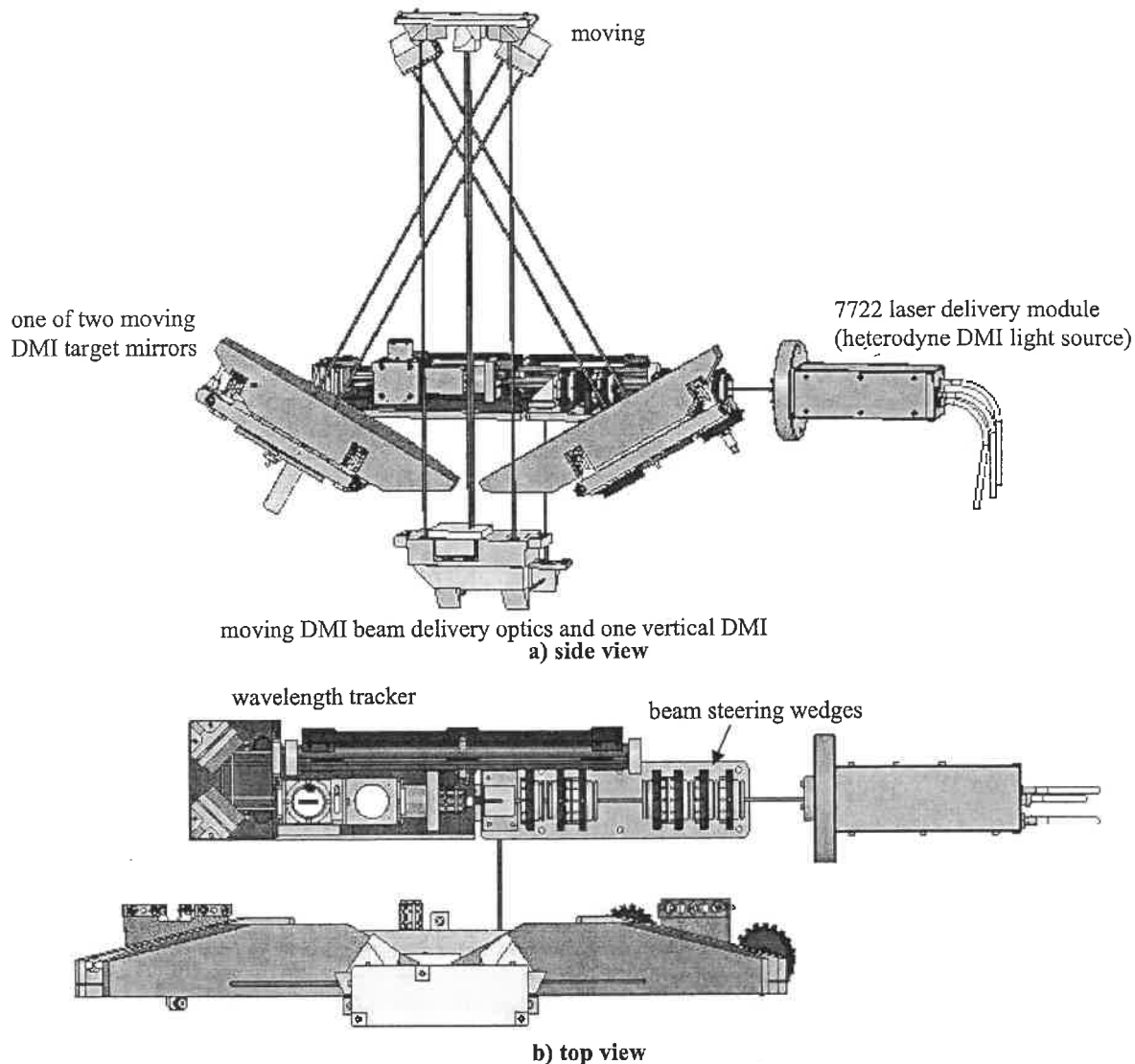


Figure 2 – 4-axis DMI system^{3,4,5}

In reality there are many sources of changing phase of the measurement signal other than target mirror displacement⁶. These other sources contribute to measurement uncertainty and include wavelength instability, beam purity, measurement arm refractive index instability, reference arm refractive index instability, interferometer beam mixing/stray light, wavefront deformation/beam shear, and phase interpolator noise.

Figure 2 shows one of the subsystems – in this case a 4-axis DMI system. The heterodyne DMI light source is a Zygo 7722 laser delivery module⁷. This fiber-fed laser module outputs more than of 1 mW of optical energy while generating less than 20 mW of heat. The output is exactly as described in the previous section with beam purity of 32 ppm (i.e., 32 parts of beam component number 1 with a polarization state matching that of beam component number 2 per 1000000 parts of beam component number 1) and a wavelength instability of <1 ppb over any 1 hour period of time (<1 part-per-billion of the nominal wavelength). The delivery module's optical output is directed to the 4-axes of DMI (3 DMIs and 1 wavelength tracker) via a set of steering wedges. Immediately following the steering wedges, 25% of the beam is allowed to pass to the wavelength tracker while the remaining 75% of the beam is directed downwards to a set of beam steering optics including two right-angle prisms and two rhomboid prisms. The rhomboids split the beam into three beams of “equal” power. One of the beams is directed into a “vertical” DMI while the other two are directed into two “moving” DMIs. Note that in this DMI system the two moving interferometers are moving with the stage motion while their target mirrors are “stationary”. The vertical DMI has a target retro-reflector which moves with the moving interferometers. All three DMIs see a change in the optical path lengths of their measurement arms with vertical motion of the stage to which they are attached (not shown in Figure 2) while only the two moving DMIs see a change in their measurement arm optical path lengths with lateral motions of the stage. With this subsystem we have redundant measurements of the vertical motions of the stage with three distinct optical paths and redundant measurements of the lateral motions with two distinct optical paths. For the purposes of this presentation/abstract, only the measurement uncertainty of one of the moving DMIs will be considered. Further, the measurement uncertainty will be considered in the coordinate frame of the interferometer, not the machine coordinate frame.

The moving DMI

The moving DMI will be operating in a vacuum (about 2 mbar) with a maximum interferometer vertical displacement of 500 nm. The measurement uncertainty is tabulated in Table 1. For the sake of brevity, only measurement uncertainties stemming from interferometer thermal sensitivity and beam sheering of non-planar wavefronts will be considered with some detail here.

uncertainty source	contribution to DMI measurement uncertainty
laser head: beam component purity	0.05 nm
wavelength instability	0.20 nm
moving DMI: thermal sensitivity	0.05 nm
beam mixing/stray light	0.10 nm
beam shear	0.60 nm
phase interpolation electronics	0.15 nm
vacuum index uncertainty	0.32 nm
Total (RSS):	0.73 nm

Table 1- Moving DMI measurement uncertainty (1 σ)

Figure 3 shows the moving interferometer. The reference arm is shown in Figure 3.b while the measurement arm is shown in 3.c. As this interferometer expands and contracts with temperature, it will grow and shrink into and out of the measurement arm some distance proportional to a length and change in temperature. The length used for this calculation is shown in Figure 3.d and is referred to as $l_{\text{sensitive}}$. What we need to do here is to calculate the temperature sensitivity of the optical path length difference between the reference and measurement arms. The optical path length difference is given by

$$\text{OPD} = n_{\text{ref}} \cdot l_{\text{ref}} - n_{\text{meas}} \cdot l_{\text{meas}}$$

where OPD is the optical path length difference between the reference and measurement arms, n_{ref} is the reference arm's index, l_{ref} is the reference arm physical length, n_{meas} is the measurement arm's index, and l_{meas} is the measurement arm physical length. Since the measurement arm consists of both glass and vacuum, it is useful to re-write the equation for optical path length difference as

$$\text{OPD} = n_{\text{SiO}_2} \cdot l_{\text{ref}} - n_{\text{SiO}_2} \cdot l_{\text{meas}_g} - n_v \cdot l_{\text{meas}_v}$$

where n_{SiO_2} is the index of the fused silica, n_v is the index of the vacuum, l_{meas_g} is the measurement arm's physical length in glass, and l_{meas_v} is the measurement arm's physical length in vacuum.

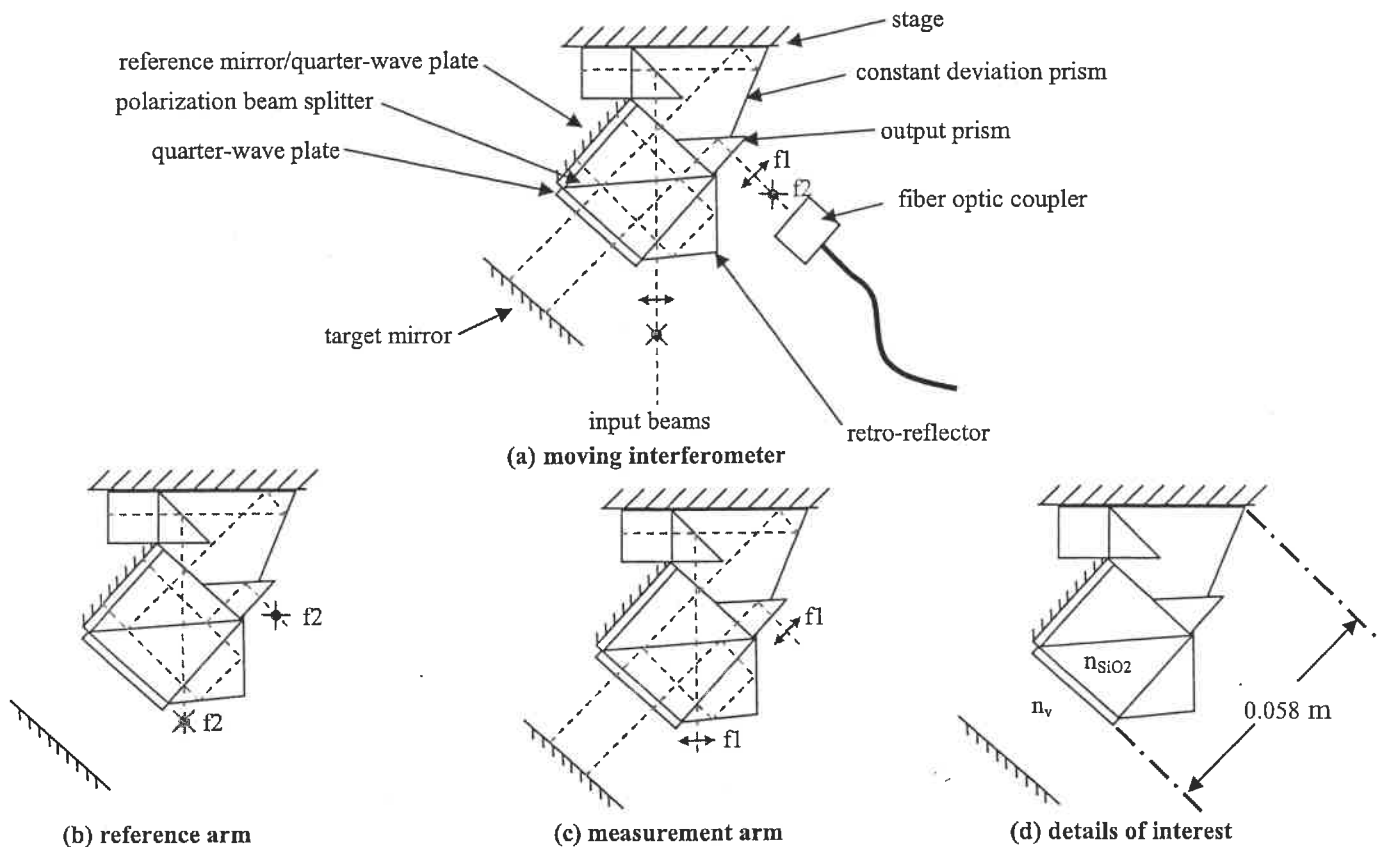


Figure 3 – Moving interferometer schematic diagrams

To find the temperature sensitivity of the interferometer we must solve

$$\frac{d(OPD)}{dT} = \frac{d(n_{SiO2} \cdot l_{ref} - n_{SiO2} \cdot l_{meas_g} - n_v \cdot l_{meas_v})}{dT}$$

$$= n_{SiO2} (l_{ref} - l_{meas_g}) \alpha_{SiO2} + (l_{ref} - l_{meas_g}) \frac{d(n_{SiO2})}{dT} - n_v (4 \cdot l_{sensitive}) \alpha_{SiO2} - l_{meas_v} \frac{d(n_v)}{dT}.$$

We ignore the last term in the equation above since measurement uncertainty due to index variations will be considered/budgeted in the analysis of the wavelength tracker. The input parameters used to solve the above equation are given in Table 2.

model input	description	expectation	standard uncertainty
n_{SiO2}	index of the fused silica	1.45698	0.000012
$d(n_{SiO2})/dT$	thermal coefficient for the index of fused silica	8.3 ppm °C ⁻¹	0.3 ppm °C ⁻¹
α_{SiO2}	thermal expansion coefficient for fused silica	0.55 ppm °C ⁻¹	0.02 ppm °C ⁻¹
n_v	index in the vacuum	1.000001327	7.78 x 10 ⁻⁹
l_{ref}	reference arm physical path length (m)	0.270 m	0.0007 m
l_{meas_g}	measurement arm physical glass path length (m)	0.270 m	0.0007 m
$l_{sensitive}$	sensitive length to thermal growth of interferometer affecting l_{meas_v} (m)	0.058 m	0.0007 m
T	temperature (°C)	23 °C	0.0014 °C
P	pressure (Torr)	3.75 Torr	0.022 Torr
F	partial pressure of the water vapor in the vacuum (Torr)	0 Torr	0 Torr

Table 2 – Model input parameter expected values and standard uncertainties

Plugging in the expected values given in the table above yields an expected interferometer temperature sensitivity of

$$X_{\frac{d(OPD)}{dT}} = 127.6 \text{ nm } ^\circ\text{C}^{-1}.$$

The corresponding standard uncertainty is

$$U_{\frac{d(OPD)}{dT}} = \left[\left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial n_{SiO2}} \right)^2 U_{n_{SiO2}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial n_v} \right)^2 U_{n_v}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial l_{ref}} \right)^2 U_{l_{ref}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial l_{meas_g}} \right)^2 U_{l_{meas_g}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial l_{sensitive}} \right)^2 U_{l_{sensitive}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial \alpha_{SiO2}} \right)^2 U_{\alpha_{SiO2}}^2 + \left(\frac{\partial(\frac{d(OPD)}{dT})}{\partial (\frac{dn_{SiO2}}{dT})} \right)^2 U_{\frac{dn_{SiO2}}{dT}}^2 \right]^{\frac{1}{2}}$$

$$= 11.5 \text{ nm } ^\circ\text{C}^{-1}.$$

The uncertainties in glass path length, vacuum path length, and the sensitive ($l_{sensitive}$) length of the interferometer dominate the standard uncertainty. The worst case optical path difference temperature sensitivity is $139.1 \text{ nm } ^\circ\text{C}^{-1}$. Note that this sensitivity has been calculated for the total optical path difference between the reference and measurement arms of the interferometer. Since there are 2 passes to the target mirror in the measurement arm, this number must be divided by 4 to get the measurement sensitivity. Next we multiply the measurement sensitivity by our temperature uncertainty to yield the resulting, conservative, moving DMI measurement uncertainty (1σ) of 0.05 nm .

Beam shear of distorted wavefronts

Some shearing of the measurement arm beam with respect to the reference arm beam within the moving interferometer will occur with changes in stage orientation. Based on the specifics of the application/design, it has been determined that this beam shear will not exceed 0.2 mm . If the measurement and reference arm beam wavefronts were planar perpendicular to the direction of their propagation, then there would be no change in the phase of the heterodyne signal as the measurement arm beam wavefront shears across the reference arm beam wavefront. However, Gaussian beam wavefronts are more spherical than planar. The wavefront departure from planar further increases as the wavefront passes through components in its optical path with inhomogeneous optical properties resulting in parts of the wavefront seeing differing indices than other parts of the wavefront thereby resulting in phase delays across the wavefront, changing its shape. As the non-planar wavefronts are sheared, there is a resulting change in phase of the heterodyne signal which is generated by mixing the electric fields on the face of a fast photodiode and whose phase is the summation of all the little phase differences between the two interfering wavefronts. For modeling purposes, we assume that all beams are Gaussian with complex amplitudes and are well represented by the equation (suppressing the time-dependence of phase)

$$E = \frac{E_0 iz_0}{z + iz_0} \exp \left[-i \frac{n2\pi}{\lambda} \left(\frac{x^2 + y^2}{2(z + iz_0)} + z \right) \right]$$

where E_0 is the electric field amplitude, Z_0 is the Raleigh distance, n is the index, $2\pi\lambda^{-1}$ is the wave number, and x , y , and z are the Cartesian coordinates. Note that the light is propagating in the z -direction. Since our beams are round in cross-section, I note that

$$x = r \cdot \cos(\theta), \quad y = r \cdot \sin(\theta), \quad \text{and} \quad r^2 = x^2 + y^2$$

where r is beam radius and θ is the θ_z . Ultimately we will multiply the sum of the measurement arm beam's electric field and the reference arm beam's electric field with its complex conjugate to get a measurement beam signal. This intensity distribution will be integrated over the active area of the detector to determine the net phase of the signal. Said another way,

$$E_{ref} = \frac{E_0 iz_0}{z_{ref} + iz_0} \exp \left[-i \frac{n_{ref} 2\pi}{\lambda_{ref}} \left(\frac{r_{ref}^2}{2(z_{ref} + iz_0)} + z_{ref} \right) \right],$$

$$E_{meas} = \frac{E_0 iz_0}{z_{meas} + iz_0} \exp \left[-i \frac{n_{meas} 2\pi}{\lambda_{meas}} \left(\frac{r_{meas}^2}{2(z_{meas} + iz_0)} + z_{meas} \right) \right], \text{ and}$$

$$I(r, \theta) = \int_0^r \int_0^\theta (E_{\text{meas}} + E_{\text{ref}})(E_{\text{meas}} + E_{\text{ref}})^* r dr d\theta.$$

We now use a phase-shifting algorithm to extract phase information from the measurement beam signal, which is proportional to the intensity described by the equation above. In this way we can predict changes in phase that result from beam shearing of the non-planar wavefronts. The magnitude of the DMI measurement error that results from beam shear is a strong function of the shapes of the wavefronts that are being sheared. One can modify the electric field equation to look like

$$E = \frac{E_0 iz_0}{z + iz_0} \exp \left[-i \frac{n2\pi}{\lambda} \left(\frac{\text{shape}(x, y)}{2(z + iz_0)} + z \right) \right]$$

where shape(x, y) is a function describing a shape. The really ambitious person could then investigate sensitivities to wavefront shape. It turns out that our wavefronts are predominantly spherical (i.e., shape(x, y) = x² + y²).

moving DMI component	number of passes	specification	expected value	uncertainty
CDP base prism	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
CDP launch	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
PBS cube	4	0.02 λ PV	0.015 λ PV	0.003 λ PV
retro-reflector	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
measurement window/reference mirror	4	0.02 λ PV	0.015 λ PV	0.003 λ PV
output prism	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
steering wedges	6	0.02 λ PV	0.015 λ PV	0.003 λ PV
right angle prism	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
Rhomboid	2	0.02 λ PV	0.015 λ PV	0.003 λ PV
fold mirror	2	0.125 λ PV	0.094 λ PV	0.018 λ PV
delivery module	1	0.02 λ PV	0.015 λ PV	0.003 λ PV
total			0.518 λ PV	0.029 λ PV

Table 3 – Moving DMI component level wavefront distortion specifications

The moving DMI component-level wavefront distortion specifications are given in Table 3. Note that the expected values for the component-level wavefront distortion specifications are assumed correlated as all of the optical components are subjected to similar manufacturing processes. Using the values in Table 3, the measurement error resulting from a 0.2 mm beam shear of the measurement arm beam with respect to the reference arm beam is 0.60 ± 0.04 nm.

Conclusions

Uncertainty analysis plays a vital role in capturing/bounding sources of uncertainty and provides the means by which a successful design can be realized. In this extended abstract, we provide an overview of our approach to reach nanometer level measurement uncertainty with heterodyne displacement measuring interferometry.

References

- 1) Zygo Corporation, "Displacement Measuring Interferometry Primer," www.zygo.com, Jan 1999.
- 2) A. Slocum, "Precision Machine Design," Prentice Hall, Englewoods Cliffs, NJ, 1992.
- 3) Henry Allen Hill, "Apparatus for generating linearly orthogonally polarized light beams," U.S. Patent No. 6157660, Dec. 5, 2000.
- 4) Van Doren, Kuechel, and Evans, "Metrology system for precision 3D motion," U.S. Patent App. 20030223078, Dec. 4, 2003.
- 5) Evans and Zandoni, "Apparatus and methods for high accuracy metrology and positioning of a body," U.S. Patent App. 20030112445, June 19, 2003.
- 6) A.E. Rosenbluth and N. Bobroff, "Optical sources of nonlinearity in heterodyne interferometers," Precision Engineering, 12, 7-11, 1990.
- 7) Zygo Corporation, "ZMI 7722 Laser Source," Operator's Manual OMP-0475A, www.zygo.com, Jan 1999.

Acknowledgements

This work is based on analytical techniques developed at Zygo Corporation over an extended period of time. Some of the key people contributing to these techniques are Dr. Henry Hill, Dr. Peter de Groot, Lars Selberg, Michael Schroeder, and Mike Metz.

Bayesian Methods for Statistical Inference on the Common Mean from Multiple Data Sources

Z.-Q. John Lu Will Guthrie

National Institute of Standards and Technology
Gaithersburg, MD

Abstract

Combining studies or data from different sources for estimation of some related or common quantities has arisen in a number of applied problems. While there are myriad results on the development of various estimators of the common mean using classical/frequentist statistical approaches, there is lack of unifying rules and understanding of the statistical inference procedures such as confidence intervals or hypothesis testing. Indeed, if the variances are different and unknown, the inferential problem is notoriously difficult, and is related to the well-known Behrens-Fisher problem. This paper will show how Bayesian statistics provides a unifying tool for uncertainty analysis in the most general multi-laboratory and multiple methods problems. Software implementations will be discussed for facilitating popularization of Bayesian methods among practical users.

1. Introduction

The problem of combining results from independent studies for the purpose of estimating or testing a common objective has been important in a number of areas, including biomedical studies and metrology. Naturally, there is a large literature in statistics and metrology that deal with this applied problem. Most of the papers in the statistical literature deal with the problem of deriving point estimators of the common mean.

Consider combining two unbiased estimators x_1 and x_2 that have mean μ and variances σ_1^2, σ_2^2 . If the ratio of the variances is known, then one can use the weighted-mean estimator

$$(1) \quad \tilde{x} = (\sigma_2^2 x_1 + \sigma_1^2 x_2) / (\sigma_1^2 + \sigma_2^2) = (x_1 / \sigma_1^2 + x_2 / \sigma_2^2) / (1/\sigma_1^2 + 1/\sigma_2^2)$$

with the associated variance

$$\text{var}(\tilde{x}) = \sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) = (1/\sigma_1^2 + 1/\sigma_2^2)^{-1}.$$

Note that both of above quantities are known if the ratio of the variances is known since the only unknown variance cancels out. If the two variances are equal, then $\tilde{x}$ reduces to the simple-mean estimator: $\bar{x} = (x_1 + x_2) / 2$. Indeed, this estimator can perform quite well even in cases where the individual variances are not equal, as its variance, $(\sigma_1^2 + \sigma_2^2) / 4$, is smaller than the variance of either individual estimator when $1/3 \leq \sigma_1^2 / \sigma_2^2 \leq 3$. In other words, it is better to combine as long as the variances do not differ too much.

In practice, of course, the variances are almost always unknown. It is precisely because the two variances need to be replaced by some data-based estimators that the problem becomes complicated. For example, Graybill and Deal (1959) show that when the variances are unknown, there does not exist any simple linear estimator of the form: $cx_1 + (1-c)x_2$ for $0 \leq c \leq 1$ which uniformly improves on the individual estimators over all possible values of σ_1^2, σ_2^2 . Thus, the weights in a linearly combined estimator have to be adaptive to the data in order to dominate over the individual estimators. It is natural to replace the unknown variances by their unbiased estimators in (1). For example, if $m_1 s_1^2 / \sigma_1^2$ is distributed as $\chi_{m_1}^2$ and $m_2 s_2^2 / \sigma_2^2$ is distributed as $\chi_{m_2}^2$, then by replacing σ_1^2, σ_2^2 by s_1^2, s_2^2 in (1), Graybill and Deal (1959) showed that the resulting combined estimator is an unbiased estimator and has smaller variance than either x_1 or x_2 when the degree of freedoms (DOFs) satisfy $m_1, m_2 \geq 9$.

In addition, in many real problems, there may be biases associated with each estimator x_1, x_2 used to estimate the common mean μ . In metrology, it is common practice to introduce a bias term, often called a type B uncertainty, in addition to the data-based variance, or type A uncertainty. For example, Eberhardt et al (1989) assumed known

bounds for the biases, and Levenson et al (2000) proposed a data-based estimation method on the bounds for the biases. The effect of introducing a bias term in the individual means is that the resulting estimator is a “shrinkage” weighted mean of the individual estimators and the related variance is inflated in order to incorporate the bias, see Section 2.

The two methods/labs problem can also be generalized to multiple methods or labs. If the number of labs or methods bigger than 5 or 10, the model presented by Rukhin and Vangel (1998), which is similar to the one-way random effects ANOVA model, but with heteroscedastic variances, is a natural framework to develop various estimators of the common mean. Rukhin and Vangel proposed a maximum likelihood estimation method, however, which is difficult to compute. They compared their estimates to various simpler approximate estimates such as the Mandel-Paul algorithm. Note that because the within-lab samples (or DOF associated with each individual mean estimator x_i) may be small in many inter-lab studies, and there will usually be many labs involved (more than 10), the shrinkage or smoothing effect on the common mean inference is more pronounced. Although the Rukhin-Vangel approach and similar approaches using random effects ANOVA as in W.G. Cochran’s approach work well in many cases, they do not address the problem of combining a small number of methods or labs.

In some senses, obtaining a point estimate of the common mean may be a relatively simple problem, since the weighted, shrinkage estimator is often close to the simple mean estimator in many cases. It is the inferential uncertainty, such as confidence intervals or hypothesis testing that appears to be the most complicated issue using the classical/frequentist approach. The reason is that, the distribution of any combined estimator of μ will involve unknown variances as nuisance parameters, so the standard methods have serious limitations in finding exact confidence intervals. For example, Jordan and Krishnamoorthy (1996) proposed some new methods for deriving exact confidence intervals and reviewed some other methods. In general, the classical frequentist statistical approach to the confidence interval is very complicated and often somewhat ad hoc. The related hypothesis testing is no less difficult.

Indeed, when the variances are unknown and unequal, the inferential problem of the common mean is related to the famous Behrens-Fisher problem. In the two-sample case, we have independent vectors $X_1 = (x_{11}, x_{12}, \dots, x_{1m})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$ such that $x_{1i} \sim N(\mu_1, \delta_1^2)$ and $x_{2i} \sim N(\mu_2, \delta_2^2)$ and the sufficient statistics for this problem is: the sample means for each vector, denoted by $\bar{x}_1, \bar{x}_2$ in accordance with previous notations, with variances $\sigma_1^2 = \delta_1^2 / m$, and $\sigma_2^2 = \delta_2^2 / n$, and sample variances for each data vector, s_1^2, s_2^2 . The Behrens-Fisher problem is the inference such as hypothesis testing on the mean difference $\Delta = \mu_1 - \mu_2$. Lee (1989), Chapter 5 gave a clear introduction to this area.

Unlike the classical frequentist statistical approaches, in which the three related problems of point estimation, confidence interval, and hypothesis testing are treated separately, the Bayesian statistical approach offers a more unified approach to the inference problems of the common mean from multiple data sets. Under the Bayesian framework, all inference issues, whether point summaries, credible or HPD intervals, or hypothesis testing, will be embodied in the posterior distribution. The difficulty of accounting for uncertainty due to estimation of nuisance parameters can be more easily dealt with through marginalization, or integration of the full posterior distribution. The Bayesian approach is well suited to problems where there are unknown and unequal variances, and where there are biases in the means that need to be estimated from data. When extra information, such as historical data or expert opinion on the method biases, it can be easily incorporated in the estimate of the common mean using Bayesian methods.

2. Bayesian Formulation of the general p Multi-lab/method Problem

Suppose that the parameters of interest from each lab/method are $\mu_1, \dots, \mu_p$. The data from p labs or methods can be summarized through data reduction using, e.g. standard statistical sufficiency theory: the statistics for the quantities of interest $\mu_1, \dots, \mu_p$ are respectively, $x_1, \dots, x_p$, with associated variances $\sigma_1^2, \dots, \sigma_p^2$. More precisely, we make the assumption that

$$(2) \quad x_1 \sim N(\mu_1, \sigma_1^2), \dots, x_p \sim N(\mu_p, \sigma_p^2), \text{ mutually independent,}$$

And that there exist Chi-squared random variables $s_1^2, \dots, s_p^2$ such that

$$(3) \quad m_1 s_1^2 / \sigma_1^2 \sim \chi_{m_1}^2, \dots, m_p s_p^2 / \sigma_p^2 \sim \chi_{m_p}^2$$

where $m_1, \dots, m_p$ are the degrees of freedom (DOFs) of the individual variance estimates from each lab. We also assume that $s_1^2, \dots, s_p^2$ are independent of $x_1, \dots, x_p$.

The parameters of interest from each lab/method $\mu_1, \dots, \mu_p$ are related to the *common* parameter of interest μ through the equations

$$(4) \quad \mu_1 = \mu + b_1, \dots, \mu_p = \mu + b_p.$$

Here the b_i 's are lab biases, which are assumed to have normal distribution with zero mean and variance s^2 .

The question now is to obtain estimates for unknown parameters of interest, $\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2$. By the Bayesian theorem, that the posterior is proportional to the product of likelihood function and the prior distribution on all unknown parameters. By the fact that the lab mean parameters may be integrated out from (2) and (4), the joint posterior distribution of the remaining unknown parameters $[\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2 | x_1, \dots, x_p, s_1^2, \dots, s_p^2]$ is proportional to:

$$(5) \quad \begin{aligned} & \propto \prod_{i=1}^p [x_i | \mu, \sigma^2, \sigma_i^2] \cdot \prod_{i=1}^p [s_i^2 | \sigma_i^2] \cdot [\mu] [\sigma^2, \dots, \sigma_p^2; \sigma^2] \\ & = \prod_{i=1}^p \exp\left[-\frac{(x_i - \mu)^2}{2(\sigma^2 + \sigma_i^2)}\right] \cdot \prod_{i=1}^p (\sigma^2 + \sigma_i^2)^{-1/2} \cdot \\ & \quad \prod_{i=1}^p (\sigma_i^2)^{-\frac{m_i}{2}} \exp\left[-\frac{m_i s_i^2}{2\sigma_i^2}\right] \cdot [\mu] \prod_{i=1}^p [\sigma_i^2] [\sigma^2 | \sigma_i^2] \end{aligned}$$

where the last term $[\mu] \prod_{i=1}^p [\sigma_i^2] [\sigma^2 | \sigma_i^2]$ is from the prior distribution assumption.

The joint posterior distribution (5) embodies all information that is needed to make inference on the unknown parameters $\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2$. For example, the Bayesian estimators of $\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2$ based on the mode of the posterior distribution are clearly related to the maximum likelihood estimators of Rukhin and Vangel (1998). But the Bayesian approach also easily handles situations when restrictions on the variance components, such as $\sigma^2 \geq 0, \sigma_1^2 > 0, \dots, \sigma_p^2 > 0$, may be in effect. More importantly, Bayesian approach offers a clear way of incorporating the uncertainty in the unknown nuisance parameters $\sigma^2, \sigma_1^2, \dots, \sigma_p^2$ in the assessment of uncertainty in the common mean inference.

Computation based on (5) is significantly simplified under balanced design assumptions, i.e. when $\sigma_1^2 = \dots = \sigma_p^2$, a natural consequence from exchangeability considerations among labs/methods. Then (5) specializes to

$$(6) \quad [\mu, \sigma^2, \sigma^2 | \text{data}] \propto (\sigma_1^2)^{-\frac{1}{2}n} (\sigma^2 + \sigma_1^2)^{-\frac{1}{2}p} \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^p m_i s_i^2}{\sigma_1^2} + \frac{\sum_{i=1}^p (x_i - \bar{x})^2}{\sigma^2 + \sigma_1^2} + \frac{p(\bar{x} - \mu)^2}{\sigma^2 + \sigma_1^2}\right]\right\} \\ \cdot [\mu] [\sigma^2 | \sigma_1^2] [\sigma_1^2], \quad -\infty \leq \mu \leq \infty, \sigma_1^2 > 0, \sigma^2 > 0,$$

where $v_1 = \sum_{i=1}^p m_i$, $\bar{x} = \frac{1}{p} \sum_{i=1}^p x_i$. To obtain the posterior distribution of the variance components (σ^2, σ_1^2) , (6) is integrated over μ yielding

$$(7) \quad [\sigma_1^2, \sigma^2 | \text{data}] \propto (\sigma_1^2)^{-\frac{1}{2}v_1} (\sigma^2 + \sigma_1^2)^{-\frac{1}{2}(p-1)} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^p m_i s_i^2}{\sigma_1^2} + \frac{(p-1)s_b^2}{\sigma^2 + \sigma_1^2} \right] \right\} [\sigma^2 | \sigma_1^2] [\sigma_1^2]$$

$$\sigma_1^2 > 0, \sigma^2 > 0,$$

Where $s_b^2 = \frac{1}{p-1} \sum_{i=1}^p (x_i - \bar{x})^2$ is an estimate of the between lab/method variability. Estimation of σ^2 and σ_1^2 should be done jointly, because of the constraints $\sigma_1^2 > 0, \sigma^2 > 0$ and the imprecise nature in the inter-lab variance statistics s_b^2 . Box and Tiao (1973, pp. 252-264) amply demonstrated this point. Overall, computation of (7) is fairly straightforward to implement, either numerically, or by Monte Carlo using χ^{-2} distribution representation as used by Box and Tiao. But, because of the heavy-tailed nature of the marginal posterior distribution of σ^2 when p is small, say ≤ 10 , then it is preferable to obtain the posterior distribution of the common mean parameter μ by direct integration of (6) over $\sigma_1^2 > 0, \sigma^2 > 0$. Alternatively, one can write the posterior of μ as an expectation of familiar densities over (7):

$$(8) \quad [\mu | \text{data}] = E_{[\sigma^2, \sigma_1^2 | \text{data}]} \Phi(\bar{x}, \delta^2) [\mu],$$

where Φ is the normal density function, and $\delta^2 = \frac{1}{p} (\sigma^2 + \sigma_1^2)$.

3. Prior Choices

For simplicity, we assume independent priors on all unknown parameters i.e. $[\mu, \sigma^2, \sigma_1^2, \dots, \sigma_p^2] = [\mu][\sigma^2] \prod_{i=1}^p [\sigma_i^2]$. When the number of observations for each method/lab is significant, estimation of individual variance components is relatively easy and is not strongly influenced by prior choices. Furthermore, estimation of μ itself is relatively stable, so we will focus on the prior choices for σ^2 mainly.

Noninformative choices are popular since they require no prior knowledge. Among the potential choices, it turns out that one has to be extremely careful about a common prior choice for variance $\pi_1(\sigma^2) = 1/\sigma^2$ or the related vague prior using inverse Gamma $\pi_1(\sigma^2) = (\sigma^2)^{-(\epsilon+1)} \exp(-\lambda/\sigma^2)$ when ϵ, λ are small. Such vague prior choice for leads to problematic issues for the posterior distribution. Indeed, such a prior choice may lead to an improper or close to improper posterior distribution. If the Laplace noninformative prior $\pi_2(\sigma^2) \equiv 1$ is used, then the posterior distribution is not proper for $p = 2$, and is only a proper density function when $p \geq 3$.

As a general recommendation, we suggest the noninformative prior:

$$(9) \quad \pi_3(\sigma^2) \propto 1/(c + \sigma^2),$$

where c is some constant to be specified, as also recommended in Box and Tiao (1973, p.372). When all labs have comparable variances: $\sigma_1^2 = \dots = \sigma_p^2 = \bar{\sigma}_a^2$, a natural choice is to take $c = \bar{\sigma}_a^2$. It can be shown that the posterior of the inter-lab variance is a proper density function, thanks in part to the partially informative prior $\pi_3(\sigma^2) \propto 1/(\bar{\sigma}_a^2 + \sigma^2)$. However, the posterior mean (and variance) of σ^2 tends to infinity when $p \leq 2$ (or $p \leq 4$) because the tail of $p_3(\sigma^2)$ decays as $(\sigma^2)^{-(p/2+1)}$ when $\sigma^2 \rightarrow \infty$. The dependence of this prior choice on the within-lab variance σ_a^2 may be reasonable, considering the fact that estimation of σ^2 , and σ_a^2 is indeed, closely correlated. Box and Tiao (1973, Chapter 3) gave a very detailed discussion of

Bayesian inference using $\pi_3(\sigma^2) \propto 1/(c + \sigma^2)$, and showed that how the Bayesian approach easily overcame some of the complications of negative estimates of σ^2 which have plagued the classical “point estimation” approaches for a long time. Though recent developments of Markov chain Monte Carlo (Robert and Casella 1999) have taken away some of the needs of clever integral manipulations in the traditional Bayesian calculations as discussed in Box and Tiao (1973), we think for the heavy-tailed posterior calculations, the direct numerical integration approach is still much needed, for calculating the posterior density function and HPD credible intervals.

Apart from the noninformative choices, one can certainly consider informative prior choices in order to model expert opinions or historical data. For example, Levenson et al (2000) has used an uniform prior on the between-lab bias. The conjugate prior choice using inverse Gamma is flexible approach for modeling the variance of between-lab bias. That is,

$$(10) \quad \sigma^2 \sim \tau^2 \chi_k^{-2}, \text{ or } IG(\tfrac{1}{2}k, \tau^2).$$

Here we use IG to denote the inverse Gamma distribution and $IG(\tfrac{1}{2}k, \tau^2)$ of degree of freedom $\tfrac{1}{2}k$ and scale τ^2 has the density function of the form:

$$\pi_4(\sigma^2) = \frac{(\tau^2)^{k/2}}{2^{k/2} \Gamma(k/2)} (\sigma^2)^{-(\frac{1}{2}k+1)} \exp\left(-\frac{\tau^2}{2\sigma^2}\right), \quad \sigma^2 > 0.$$

The general setup is easily amenable to a Markov Chain Monte Carlo implementation through (5). To obtain the marginal distribution of μ, σ^2 , one can simply draw random samples of $\sigma_1^2, \dots, \sigma_p^2$ from inverse chi-squares, then use Gibbs sampling to generate (dependent) samples through μ , and σ^2 alternatively.

4. Examples

The proposed Bayesian methods have been applied to some typical common mean inference problems in the literature, including the Eberhardt et al (1989) example, also studied by Jordan and Krishnamoorthy (1996). Another example is a non-Gaussian example, in which one needs to combine binomial trials using a Bayesian Beta-binomial model in order to account for the between-sample variability.

References

- [1] Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Inference*. Addison-Wesley, Reading, Massachusetts.
- [2] Eberhardt, K.R., Reeve, C.P. and Spiegelman, C.H. (1989). A Minimax Approach to Combining Means, With Practical Examples. *Chemometrics and Intelligent Laboratory Systems*, **5**, 129-148.
- [3] Graybill, F.A. and Deal, R.B. (1959). Combining Unbiased Estimators. *Biometrics*, **15**, 543-550.
- [4] Jordan, S.M. and K. Krishnamoorthy (1996). Exact Confidence Intervals for the Common Mean of Several Normal Populations. *Biometrics*, **52**, 77-86.
- [5] Lee, P. M. (1989). *Bayesian Statistics: An Introduction*. Oxford University Press, New York.
- [6] Levenson, M.S., D.L. Banks, K.R. Eberhardt, L.M. Gill, W.F. Guthrie, H.K. Liu, M.G. Vangel, J.M. Yen, and N.F. Zhang (2000). An Approach to Combining Results from Multiple Methods Motivated by the ISO GUM. *J. Res. Natl. Inst. Stand. Technol.*, **105**, No.4, 571-579.
- [7] Rukhin, A.L. and M.G. Vangel (1998). Estimation of a Common Mean and Weight Means Statistics. *Journal of American Statist. Association*, **93**, No.441, 303-308.

Uncertainty Analysis of Interlaboratory Studies with Linear Trends

Nien F. Zhang, Hung-kung Liu, Nell Sedransk, and William E. Strawderman
National Institute of Standards and Technology, Gaithersburg, MD

Interlaboratory studies have been used to ensure accurate measurement capability among different laboratories. In this paper, we discuss a special interlaboratory comparison called key comparison. With the recent signing of the Mutual Recognition Arrangement (MRA), national metrology institutes (NMIs) and regional metrology organizations (RMOs) around the world have committed themselves to establishing the equivalence of their measurement standards through key comparisons of national measurement standards. In most key comparisons, the pilot NMI is responsible to organizing the circulation and transport of standards or artifacts and ensuring that the participant NMIs make proper arrangements. See the Appendix F of the “Mutual Recognition of National Measurements Standards and of Measurement Certificates issued by National Metrology Institutes” [1]. Usually each non-pilot NMI makes measurements in one period while the pilot NMI makes measurements in multiple periods based on the circulation scheme. In some key comparisons, the measurements of the transport standards made by the participating NMIs will show a trend or a drift (a linear drift in this paper). If this occurs, a traditional statistical analysis, which assumes no trend exists as in the recent publications, is inappropriate. Within some consultative committees (CCs) of the International Committee for Weights and Measures (CIPM), some uncertainty analyses are performed to accommodate the drift effect. One approach is for each NMI to calculate the difference between the measured value and the predicted value based on the trend, e.g., a linear trend. By combining these differences, e.g., using a weighted mean, the key comparison reference value (KCRV) is then calculated. However, in the process of calculating the uncertainties relating to the KCRV and other quantities, the correlations among the predicted values of the non-pilot NMIs, which are jointly based on the linear regression from the measured values of the pilot NMI, were not accounted for and thus a more rigorous statistical study was needed. In the paper by Zhang, Sedransk, and Jarrett [2], an approach was proposed and applied to the CCEM-K2 Key Comparison of Resistance Standards at 10 M Ω and 1 G Ω (see Dziuba and Jarrett [3]). The proposed method in the above article improved the uncertainty analysis by considering the correlations (which were ignored in many practices) due to the predictions for non-pilot laboratories jointly based on the pilot laboratory measurements. However, due to the definition of KCRV in the existing methods dealing with trend, the results cannot be applied to the trivial case when the trend reduces to zero. In this article, we define the time-dependent KCRV in an appropriate way and propose a new approach that overcomes the shortcomings in [2] and [3]. The calculation of KCRV is consistent with the case in which there is no trend. The pair-wise comparison between any two national measurement standards and its corresponding uncertainty, which is independent of the time, are also given.

We assume that the measurements of any particular laboratory have a linear trend in time and the slopes of the linear trends for all the laboratories are the same while we allow for different intercepts for different laboratories. In this article we consider the case of a

single artifact circulating through all laboratories. Without loss of generality, we assume that the pilot laboratory is the first one among all I laboratories. Denote the time and the result of the k^{th} measurement by the pilot laboratory by t_{1k} , $k = 1, \dots, K$, which are taken to have negligible associated uncertainty, and X_{1k} , $k = 1, \dots, K$ with $K > 2$, respectively. In practice, t_{1k} can be an average value of the time when the measurements were made in the k^{th} period and then X_{1k} is the average value of the corresponding measurements in that period.

We assume that a simple linear regression model holds for the measurements made by the pilot laboratory,

$$X_{1k} = \alpha_1 + \beta t_{1k} + \varepsilon_{1k} \quad (1)$$

for $k = 1, \dots, K$. We assume that the random error ε_{1k} has a zero mean and uncertainty of u_1 . When $i \neq 1$, each laboratory takes one measurement at time t_i and the corresponding model is

$$X_i = \alpha_i + \beta t_i + \varepsilon_i, \quad (2)$$

where the random error ε_i has a zero mean and standard uncertainty of u_i for $i = 2, \dots, I$. In a key comparison study, NMI's will provide x_i 's, the values of X_i 's, and the corresponding uncertainties including the Type A and Type B evaluations of uncertainty.

For the pilot laboratory, the least squares estimators of the regression parameters are given by

$$\hat{\beta} = \frac{\sum_{k=1}^K (t_{1k} - t_1)(X_{1k} - X_1)}{\sum_{k=1}^K (t_{1k} - t_1)^2} \quad (3)$$

and

$$\hat{\alpha}_1 = X_1 - \hat{\beta} t_1, \quad (4)$$

where X_1 is the average of the measurements made by the pilot laboratory and t_1 is the average of $\{t_{11}, \dots, t_{1K}\}$. For other laboratories, i.e., $i = 2, \dots, I$, the corresponding intercepts can be estimated by

$$\hat{\alpha}_i = X_i - \hat{\beta} t_i, \quad (5)$$

where t_i is the time when the i^{th} ($i \neq 1$) laboratory made its measurement.

Key Comparison Reference Value

One approach to calculating the KCRV at any time t (denoted by $KCRV_t$) is to use a weighted mean of $\hat{\alpha}_i + \hat{\beta}t$ over the laboratories $i = 1, \dots, I$,

$$KCRV_t(w) = \sum_{i=1}^I w_i(\hat{\alpha}_i + \hat{\beta}t) = \sum_{i=1}^I w_i X_i - \hat{\beta} \sum_{i=1}^I w_i(t_i - t) \quad (6)$$

with weights $w = (w_1, \dots, w_I)$ satisfying $\sum_{i=1}^I w_i = 1$. $KCRV_t(w)$ is a function of t as well as the weights. The uncertainty of KCRV depends on t as well as the weights w . For any given t , a reasonable criterion to choose optimal weights is the minimum variance or minimum standard uncertainty of $KCRV_t(w)$. For any given set of weights, the minimum value of the standard uncertainty of $KCRV_t(w)$ will occur when $\sum_{i=1}^I w_i(t_i - t)^2$ equals zero. This occurs when t takes the value of $t_w = \sum_{i=1}^I w_i t_i$. In this case, from (6), the KCRV has the same form as in the no trend case and is given by

$$KCRV_{t_w}(w) = \sum_{i=1}^I w_i X_i. \quad (7)$$

The corresponding uncertainty is given by

$$u_{KCRV_{t_w}(w)}^2 = \sum_{i=1}^I w_i^2 u_i^2. \quad (8)$$

As in the no trend case, $u_{KCRV_{t_w}(w)}^2$ is minimized when the weights are given by

$$w_i(1) = \frac{\frac{1}{u_i^2}}{\sum_{k=1}^I \frac{1}{u_k^2}}. \quad (9)$$

The corresponding KCRV is given by

$$KCRV_{t^*}(w(1)) = \sum_{i=1}^I w_i(1) X_i, \quad (10)$$

where

$$t^* = \sum_{i=1}^I w_i(1) t_i, \quad (11)$$

and $w(1) = (w_1(1), \dots, w_I(1))$. From (9) and (10), the uncertainty of this KCRV is the same as in the no-trend case and is given by

$$u_{KCRV_{t^*}(w(1))}^2 = \frac{1}{\sum_{i=1}^I \frac{1}{u_i^2}} \quad (12)$$

assuming that u_i ($i = 1, \dots, I$) in the weights $\{w_i(1)\}$ are the standard deviations for all laboratories. As a practical matter, we recommend the $KCRV_{t^*}(w(1))$ with t^* and the weights given by (9) and (11) to be the reference value of this key comparison.

Degrees of equivalence of the national measurement standards with respect to the KCRV

The degree of equivalence of the national measurement standard from the i^{th} laboratory with respect to the $KCRV_t$ is defined as the difference:

$$\begin{aligned} D_{iKCRV(w)} &= \hat{\alpha}_i + \hat{\beta}t - KCRV_t(w) \\ &= (1 - w_i)\hat{\alpha}_i + \sum_{j \neq i, j=1}^I w_j \hat{\alpha}_j. \end{aligned} \quad (13)$$

when $w = w(1)$ as given in (9), in the first equality of (13), let $t = t^* = \sum_{i=1}^I w_i(1)t_i$. Then

$$\begin{aligned} D_{iKCRV(w(1))} &= \hat{\alpha}_i + \hat{\beta}t^* - KCRV_{t^*}(w(1)) \\ &= \hat{\alpha}_i + \hat{\beta}t^* - \sum_{i=1}^I w_i(1)X_i. \end{aligned} \quad (14)$$

The corresponding uncertainty is given by

$$u_{D_{iKCRV(w(1))}}^2 = [1 - 2w_i(1)]u_i^2 + \frac{1}{\sum_{k=1}^I \frac{1}{u_k^2}} + \frac{(t_i - t^*)^2 \sigma_{1,A}^2}{\sum_{k=1}^K (t_{1k} - t_1)^2}, \quad (15)$$

where $\sigma_{1,A}$ is the Type A evaluation of uncertainty of the measurements made by the pilot laboratory. An unbiased estimator of $\sigma_{1,A}^2$ in (15) is given by the “mean-square residual” based on the model in (1), i.e.,

$$\hat{\sigma}_{1,A}^2 = \frac{\sum_{k=1}^K (X_{1k} - \hat{\alpha}_1 - \hat{\beta}t_{1k})^2}{K - 2}. \quad (16)$$

Again we recommend using $D_{iKCRV(w(1))}$ as the degree of equivalence of the national measurement standard from the i^{th} laboratory with respect to the KCRV.

Degrees of equivalence of pairs of national measurement standards

The degree of equivalence between the national measurement standards at time t is defined as

$$D_{i,k} = (\hat{\alpha}_i + \hat{\beta}t) - (\hat{\alpha}_k + \hat{\beta}t) = \hat{\alpha}_i - \hat{\alpha}_k \quad (17)$$

when $i \neq k$. Thus, the quantity is independent of t . The corresponding uncertainty is

$$u_{D_{i,k}}^2 = u_i^2 + u_k^2 + \frac{(t_i - t_k)^2 \sigma_{1,A}^2}{\sum_{k=1}^K (t_{1k} - t_1)^2}. \quad (18)$$

To illustrate the approach, we applied it to the key comparison CCEM-K2. In this key comparison, three 10 M Ω wirewound resistors and three 1 Ω film-type resistors were used as the transport standards. During the comparison, the transport standards were measured at the pilot NMI, NIST. For each period an average value of the dates when the measurements were made is called a mean date of the period. Each non-pilot NMI reported a mean value and a mean date of measurement for each of the six artifacts. An uncertainty budget that includes the Type A and Type B evaluations of uncertainties for each NMI's measurement process was also reported. In this example, we only consider the 10 M Ω case and use the measurement data only for the artifact, S/N HR7551 listed in Table 1. The uncertainties for all NMIs are also listed in Table 1.

For the trend, the assumption is that the resistor drifts in a linear fashion. Any non-linear effects are caused probably by several physical or mechanical changes during the transportation process. A linear regression line was fit to the NIST measurements. The estimates of the intercept for NIST, the slope, and the residual standard deviation are $\hat{\alpha}_1 = 6.03$, $\hat{\beta} = 1.05$, and $\hat{\sigma}_{1,A} = 1.09$, respectively. We used $\hat{\sigma}_{1,A}$, the residual standard deviation from (16) as an estimate of $\sigma_{1,A}$ in (15) and (18). The date for minimum standard deviation of KCRV is $t^* = 1998.23$ given by (11). The KCRV at t^* with weights $w(1)$ given by (9) and (10) is 8.03 and the corresponding uncertainty is 0.28 given by (12). The $D_{iKCRV_t^*(w(1))}$ for all NMIs and their corresponding uncertainties $u_{D_{iKCRV_t^*(w(1))}}$ were calculated.

The degrees of equivalence of pairs of NMIs and their corresponding uncertainties are calculated by (17) and (18), respectively.

In summary, in this paper we propose a new statistical analysis for key comparisons with linear trends. The calculation of KCRV is consistent with the case in which there is no

trend. The corresponding uncertainties for KCRV and degrees of equivalence are also provided.

1. Guidelines for CIPM key comparisons (Appendix F to *Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes*.) Technical Report, International Committee for Weights and Measures, 1999.

2. Zhang, N. F., Sedransk, N., and Jarrett, D. G., Statistical Uncertainty Analysis of Key Comparisons CCEM-K2, *IEEE Transaction Instrumentation & Measurement*, 2003, 491-494.

3. Dziuba, R. F. and Jarrett, D. G. Final Report CCEM-K2 Key Comparison of Resistance Standards at 10 M Ω and 1G Ω . Appendix B of *Mutual Recognition Arrangement*, 2002, Bureau International des poids et Mesures (BIPM), Paris France.

Table 1. Information for the Standard S/N HR7551

Lab	Mean date of Measurement	Measurement (x 10 ⁻⁶)	Type A uncertainty (x 10 ⁻⁶)	Type B uncertainty (x 10 ⁻⁶)
NIST	1996.65	4.6	0.2 *	1.51*
NRC	1996.80	5	1.88	2.29
NIST	1996.94	6.7	0.2 *	1.51*
BNM-LCIE	1997.17	6.97	0.50	0.35
NPL	1997.35	7.1	0.52	0.61
PTB	1997.50	7.5	1.0	2.19
NIST	1997.62	8.1	0.2 *	1.51*
CSIRO-NML	1997.82	7.3	0.07	2.56
MSL	1998.03	7.3	0.04	0.59
CSIR-NML	1998.13	-20	50	13.52
NIST	1998.33	8.9	0.2 *	1.51*
SP	1998.49	8.7	0.17	1.79
OFMET	1998.62	8.9	0.39	0.58
IEN	1998.74	9.4	0.79	2.53
NMi-VSL	1998.98	9.1	0.80	3.04
NIST	1999.15	7.8	0.2 *	1.51*
KRIST	1999.39	7.1	0.30	3
NIST	1999.60	8.5	0.2 *	1.51*
NIM	1999.87	10.2	0.10	0.83
VNIIM	2000.03	10	0.25	1.03
NIST	2000.20	10	0.2 *	1.51*

Uncertainty Analysis of Kolsky Bar Strain Gage Output

Richard Rhorer
National Institute of Standards and Technology
Gaithersburg, MD 20899-8223

An apparatus has been built at the National Institute of Standards and Technology (NIST) to measure the dynamic stress-strain relationships for metals at elevated temperatures. The NIST Pulse Heated Kolsky Bar apparatus has the capability to resistively heat samples to over 1000 C with heating rates of over 10 000 C/s in addition to performing the traditional Kolsky bar impact testing. The traditional Kolsky bar (or Split Hopkinson Pressure Bar) consists of two long hardened steel bars aligned end-to-end with the sample sandwiched in between the bars. A test is carried out by striking the end of the incident bar with a projectile fired from an air gun that sends an elastic wave down the bar, which then rapidly deforms the sample. As the sample is deformed a transmitted wave is started down the second bar. Standard variable resistance metal foil strain gages are mounted in the center of each of the long bars that measure the amplitude of the strain wave as a function of time. The outputs of the strain gages are recorded on a standard digital oscilloscope data acquisition system. Using the strain gage signal recordings a high-strain rate stress-strain curve for the sample material can then be calculated using a well established method. Our Kolsky bar apparatus can test samples at strain rates from about 500 /s to 10 000 /s (normal compression machines in materials test labs operate in the range of 0.01 /s and under special conditions can go as high as 100 /s.) By measuring the temperature of the sample during the deformation (the deformation takes approximately 150 microseconds) the stress-strain curve can be reported for various temperatures.

The strain gage measuring system is calibrated by a standard method of shorting a large precision resistor across the gage and recording the output. The uncertainty of the simulated strain in this parallel resistor calibration method depends on the uncertainty of the value of the parallel resistor, the gage resistance, and the manufacturer's gage factor. There is an additional contribution to the uncertainty of the measurement results because of the dynamic effects on the gage factor. To understand the accuracy of the strain gage output the different uncertainties need to be combined in an uncertainty analysis. The uncertainty analysis of the strain gage calibration method presented in this paper uses an approach based on NIST Technical Note 1297. Future papers will address the uncertainty of the temperature measuring methods and some of the effects of the assumptions used in the Kolsky bar simple theory for relating strain in the elastic bars to the stress and strain in the sample.

Characterization of a Turbine Airfoil Cooling Hole Check Standard for 5-axis CMMs - Uncertainty Reduction by Redundancy and Error Reversal

Joe Drescher, UTC - Pratt & Whitney, East Hartford, CT

Abstract

Measurement of the location of cooling holes generally requires a 5-axis, dual-sensor system. The size of the holes dictates an optical measurement while a touch probe is best to establish the location and orientation of the part. The measurement requires tilt and rotary axes to position the nominal hole axis parallel to the optical axis. Methods to certify volumetric performance of 5-axis, dual-sensor CMMs are not well established. One approach is the use of a check standard. This paper describes a sensitivity analysis approach to the estimate of task specific measurement uncertainty for the positions of holes on a check Standard. The mathematical model for the measurement is more complicated than is usually solved by this method [1].

1. Introduction

A check standard is to be used periodically as part of CMM calibration [2]. It has to be representative of all blades and stators to be inspected in terms of cooling hole size and configuration such that the full range of each machine axis is used in the check standard inspection. It was designed to have simple, prismatic geometry so that its coordinate system (CS) could be established accurately. Also, the holes were specified slightly larger than those in actual blades so that they could be produced with better form, thus reducing the measurement uncertainty related to this. Surface finish and optical properties were also targeted to match those of airfoil castings.

After the check standard was made, the first step was to measure the actual position of each hole. The uncertainty for the measurement had to be less than the capabilities of available 5-axis CMMs, that is, the machines to be tested. Traditional, bench layout was considered inadequate for all but the simplest hole location and orientation. An existing 5-axis CMM was chosen for the task with the expectation that special measurement methods could be used to improve the accuracy of results. The machine employs a standard touch-trigger probe to locate the part and a triangulation laser to locate each hole by edge detection. Multiple measurements and reversal techniques were used to reduce non-systematic and systematic components of measurement uncertainty, respectively, beyond what could have been achieved in a single measurement. Measurement uncertainty was estimated in two orthogonal directions for each of 44 holes. The 95% confidence interval ranged from a low value of 22.3 μ m to a high value of 36.8 μ m.

2. Measurement Method

The orientation of an airfoil in the hole measuring position is not unique. In general, there are two symmetrical and opposite orientations; clockwise (CW) and counterclockwise (CCW), corresponding to the tilt axis direction. The hole is measured as x-y true position in the hole CS defined such that Z is parallel to the nominal hole centerline.

Each measurement includes in each direction:

- The actual deviation of the hole position from nominal.
- Non-systematic error specific to CW or CCW orientation.
- Systematic errors specific to CW or CCW orientation.
- Systematic errors common to CW and CCW orientation.

The last group of errors is substantial in the dual-probe, 5-axis system. They include

- Datum of the measurement probe in the machine CS.
- Zero setting of linear axes.
- Zero setting of rotary and tilt axes.
- Coincidence error of two rotational average axes of rotation.
- Perpendicularity of Z axis to XY measurement plane.
- Parallelism of tilt axis to XY measurement plane.
- Perpendicularity of rotary and tilt axes of rotation.
- Various parametric errors of axes that do not move from CW to CWW orientation.

These were eliminated from the measurement error by the decision to measure both CW and CCW and calculate the hole position as half the difference of the positions measured in each direction. The actual deviation from nominal switches sign along with the part CS as it is rotated by 180 degrees within the machine CS while these bias terms do not switch sign.

The same errors, however, also occur during probing of the part to establish the part CS with respect to machine CS and therefore are not completely eliminated from the measurement results. They switch sign along with the measurand from CW to CCW during subsequent measurement. For the check standard, a simple change to the probing algorithm was adopted. It was probed in two mirror image sequences and the results of displacements and rotations of the two determined part CS were averaged. This eliminated the systematic errors that were common to both mirror image positions during part probing.

3. Analysis

Error in measurement, ε , then included all nonsystematic errors lumped together, NSE , the systematic errors of probing after reversal, PE' , the systematic errors of hole measurement after reversal, ME' , and the errors due to laser probe, part, sampling method, fitting and the interactions, $PPSFE$.

$$\varepsilon = NSE + PE' + ME' + PPSFE \quad (1)$$

The four terms were assumed linearly independent. That is, the components of error that would cause PE and ME to be correlated (the bias terms listed above) were simply neglected in the definitions of PE' and ME' . The expression for combined standard uncertainty is:

$$u^2(\varepsilon) = u^2(NSE) + u^2(PE') + u^2(ME') + u^2(PPSFE) \quad (2)$$

3.1 Standard Uncertainty due to nonsystematic errors

The nonsystematic error was estimated statistically from 10 measurement sets with $n = 10$, and estimate of standard deviation, s .

$$u^2(NSE) = s^2 / n$$

The standard uncertainty due to nonsystematic errors was calculated to be from 0.41 μm to 4.3 μm .

3.2 Standard Uncertainty due to errors during probing

For the final probing iteration, uncertainty was estimated for each measured deviation contained in P , a 1X6 matrix of the probe deviations measured from the nominal, expected location. The check standard is shown in Figure 1 with the 6 probe points and their surface normal vectors numbered 1-6. It is a simple plane (1-3), line (4,5), point (6) system.

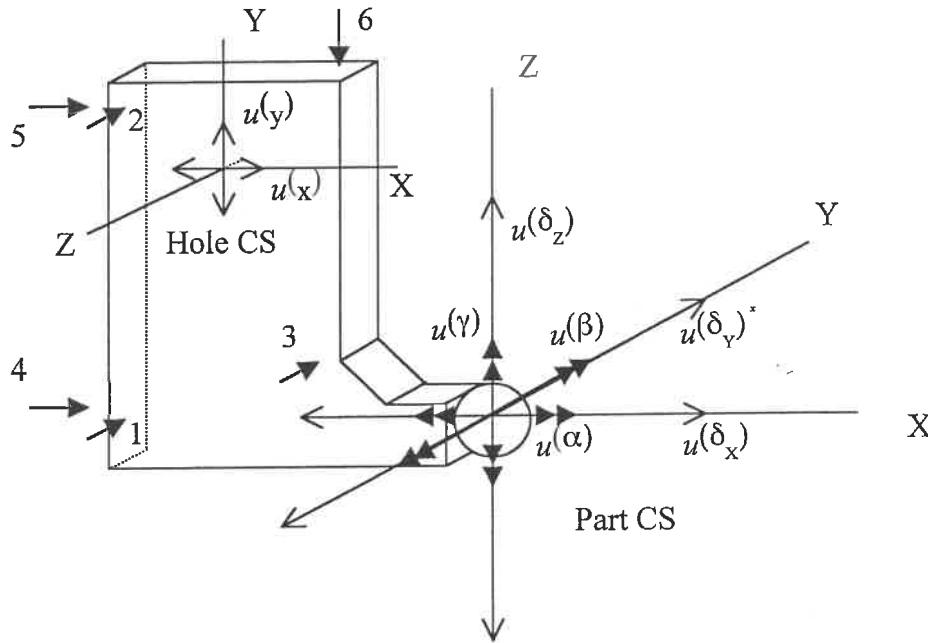


Fig. 1. The check standard artifact schematic showing probe points, nominal part CS, and deviations of part CS from nominal.

The standard uncertainty of each probe deviation was estimated using guidelines for estimating uncertainty [3]. Unknown, systematic machine errors were modeled with uniform probability and limits as measured during machine calibration. Appropriate fixed offsets were included taking into account the part geometry and location of the probe point relative to lines of measurement in the pre-process metrology. The list of unknown, systematic errors excluded those common to the symmetrical probing orientation. So, for example:

$$u^2(P_j) = \frac{n}{3} \sum_{i=1}^n \left(\frac{R_i}{2} OF_i \right)^2 \quad (3)$$

where:

R_i was the range of each error from the initial metrology.

OF_i was the offset factor including unit conversion.

The offset factor was nonzero for X-Y squareness, A-axis angular accuracy and Y-axis pitch for the 1st probe point in the Y direction. Other error parameters included Y-axis linear accuracy and X-axis straightness. Equation (3) was applied to each of the 12 probe points, 6 in the cw and 6 in the ccw orientation.

Six parameters describing the deviation from nominal of the part CS are functions of the six probe deviations, P_1, P_2, P_3, P_4, P_5 , and P_6 contained in $\mathbf{P}$ for both cw and ccw probe sequences. The three translations and three rotations may be written as $\mathbf{D} = [\delta x \ \delta y \ \delta z \ \alpha \ \beta \ \gamma]^T$. Uncertainties for these parameters were estimated by rules for propagation of uncertainty using the coefficient matrix, $\mathbf{C}$ 6X6, where,

$$\mathbf{D} = \mathbf{C} (\mathbf{P}_{cw} + \mathbf{P}_{ccw}) \quad (4)$$

$\mathbf{C}$ is comprised of part coordinates of the probe points. For example, angular deviation, α , of the part CS about the nominal X direction in expanded form is given by:

$$\alpha = \frac{1}{2} \left[\frac{1}{Z_2 - Z_1} (P_{1cw} + P_{1ccw}) - \frac{1}{Z_2 - Z_1} (P_{2cw} + P_{2ccw}) \right] \quad (5)$$

Clearly the errors in measured deviations at probe points may be correlated. Correlation coefficients were established by Monte Carlo simulation. Ranges of machine errors from pre-process metrology were taken as limits on uniform probability distributions with proper offsets applied to angular terms. Many sets of errors were generated to estimate possible probing error at each point by summation of individual errors and the correlation coefficient was calculated using the spreadsheet function. Thus the standard uncertainty was estimated for each parameter describing the location and orientation of the part CS.

The six uncertainty terms associated with the part CS were then transformed into each hole CS. The standard uncertainties of translation of the part CS origin were transformed directly into the hole CS by the two appropriate rotations for each hole. The standard uncertainties of orientation of the part CS were individually used to first calculate a translational uncertainty at the hole location in the part CS. This was then transformed into the hole CS through the appropriate angles. The standard uncertainty due to probing in X and Y directions of the hole CS as shown in Figure 1 was then the combination of the six terms taken as independent. This value ranged from 1.8 μm to 7.5 μm .

3.3 Standard Uncertainty due to errors during measurement

The reversal method, i.e. taking the measurement results as half the difference between CW and CCW results, eliminated many systematic errors. The remaining systematic errors as measured in the initial machine calibration were used to calculate the measurement uncertainty at each hole location. There was assumed to be no correlation between CW and CCW errors for any holes. In all but a few instances, tilt, rotary, X and Y axes all change position between the two mirror image orientations. Therefore, the standard uncertainty in each direction for each hole was taken as the combination of standard uncertainties of

individual machine errors. Again, these were modeled as having uniform probability distributions with bounds equal to the measured ranges. The standard uncertainty due to errors during measurement ranged from a low estimate of $1.6\mu\text{m}$ to a maximum of $3.1\mu\text{m}$.

3.4 Part, probe, sampling, fitting interactions

A laser is focused to a nominal $80\mu\text{m}$ spot and this is positioned to the nominal center of the hole. A preliminary 4 point edge finding sequence adjusts the center. The measurement then consists of 6 radial moves from the new center at 60 degree increments. A threshold intensity on the central area of the ccd array triggers a capture of the machine position at the hole edge. A least squares fit to the six points yields the measured center. Error sources include:

1. Edge condition such as debris and burrs
2. Velocity / effective radius / center start effects
3. CCD resolution
4. Laser shape at focus
5. Laser / detector asymmetry
6. Actual angle of the centerline of the cylinder (hole)
7. Hole roundness
8. Finite sampling

Items 1, 2, and 3 may be considered nonsystematic effects. They contribute to repeatability and are included in the NSE term. The shape of the laser was observed to have a symmetrical, elliptical cross section aligned with the machine axes. Therefore, because the primary measurement is the hole center and not size, this error was not included in the analysis. The laser detector asymmetry and angle of the hole were assumed to be eliminated by the measurement of each hole at 0 and 180 degrees (Reversed). The asymmetry of the single, 25degree detector angle may result in a smaller effective radius as the probe moves in the direction away from the detector. Light may be captured returning from the ID of the cylinder wall. The effect should be captured to some extent in the probe calibration and also it should be canceled in the reversal because the measurement is center position, not hole size. A similar effect may be caused when the angle of the cylinder varies from nominal and a similar argument may be made that this effect should be eliminated by the reversal.

The form error of the holes was observed to be significant considering the finite sampling of 6 edge positions to determine the center. Nine holes representing various positions and angles were measured using a modified probe routine with 60 points. The least squares center of the higher resolution edge profile was compared to the center using only the six points. A standard deviation was calculated from the 18 differences (9 x values and 9 y values) and this was used as the standard uncertainty due to part, probe, sampling, and fitting on all holes. $u(\text{PPSFE}) = 4.9\mu\text{m}$

4 Results and verification

The combined standard uncertainty from equation (2) with a coverage factor of 2 was used for the error bars about the average measurement. This is plotted in Figure 2. The maximum expanded uncertainty is $36\mu\text{m}$ and the minimum is $22\mu\text{m}$. The check standard artifact was used to test three competing CMMs for a new equipment order. Some of the holes were

also checked by bench layout techniques. These four sets of independent data are also included in Figure 2. All data is normalized to the calibration results.

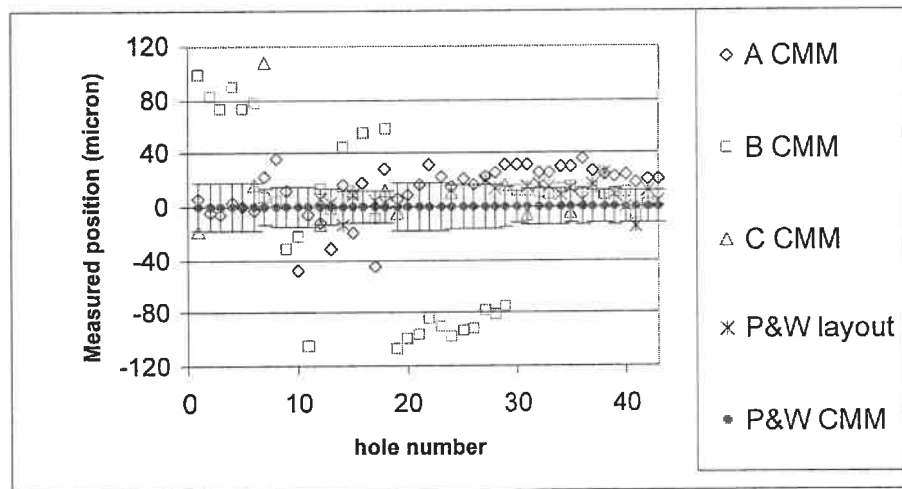


Fig. 2. Measurement results with expanded uncertainty shown by error bars and compared to four independent sets of measurements. P&W CMM measurements are set to zero and all others are shown as deviations.

5. Discussion

A few things should be analyzed to refine the uncertainty analysis. Treatment of the unknown, systematic squareness terms as uniformly distributed between specified bounds is not conservative. If the actual value is near one limit as is likely, the magnitude of the covariance term calculated by the simulation is too low. If squareness is large compared to the other errors, the CW and CCW results would be more highly correlated than estimated here. The best method is probably to correct the measurements for squareness although this is not trivial considering the part probing and measurement methods used.

Roll of X-axis and Y-axis were considered negligible as all measurement points were within 25mm of the line of straightness measurement. These terms may be added to a refined analysis.

The possible interactions between the laser probe system and the condition of the hole edges could be analyzed more fully. There are significant differences in the sensitivity to this between the laser probing method, optical contrast imaging and analysis, and manual measurement of pins.

References

- [1] Wilhelm R.G, Hocken R, Schwenke, H. Task Specific Uncertainty in Coordinate Measurement. Ann CIRP 2001;50/2:553-63
- [2] NCSL Z540.1. Calibration laboratories and measuring and test equipment. 1994
- [3] ISO Guide. Guide to the expression of uncertainty in measurement. ISBN 1993;92-67-10188-9

Using Homogeneous Transformation Matrices to Define the Measurand for Uncertainty Analysis of Complex Measurement Systems

T. Schmitz
J. Ziegert
J. Suzanne Canning

Machine Tool Research Center
University of Florida
Gainesville, FL 32611

ABSTRACT

In this paper, we consider the problem of uncertainty analysis of complex measurement systems with multiple axis motions and a measurand defined by non-linear combinations of input quantities. We demonstrate a method for obtaining the definition of the measurand using Homogeneous Transformation Matrices (HTM) to model the instrument and obtain the measurand as a function of the motion and sensing errors of the individual axes. The method is demonstrated by application to the laser ball bar (LBB), which is capable of measurement of spatial coordinates of a target over relatively large working volumes approaching 1 m^3 .

1.0 Introduction

The first step in performing measurement uncertainty analyses is the definition of the measurand, or quantity being measured. In many cases, the measurand is not observed directly but is a mathematical function of other input quantities. The computation of measurement uncertainty from the uncertainties of the input quantities is then a relatively straightforward procedure and is described in ANSI/NCSL Z540-2-1997, the US Guide to the Expression of Uncertainty in Measurement [1], and the ISO GUM [2].

It is well recognized that "incomplete definition of the measurand" can produce significant sources of measurement error and uncertainty. In general, as the measurement instrument or system becomes more complex, more input quantities will be present and correct definition of the measurand becomes more problematic. This paper describes the use of Homogeneous Transformation Matrices (HTM) as a tool to aid in defining the measurand and performing uncertainty analyses for the laser ball bar, a complex, multi-axis measurement system. This approach follows the initial HTM measurand definition for optical bench radius measurements described by Davies and Schmitz [3].

2.0 Laser Ball Bar

The laser ball bar (LBB) is an instrument designed to evaluate the positioning accuracy of machine tools by direct measurement of tool point coordinates relative to the work surface over the working volume of the machine [4,5,6,7]. The working principle of the laser ball bar is described below.

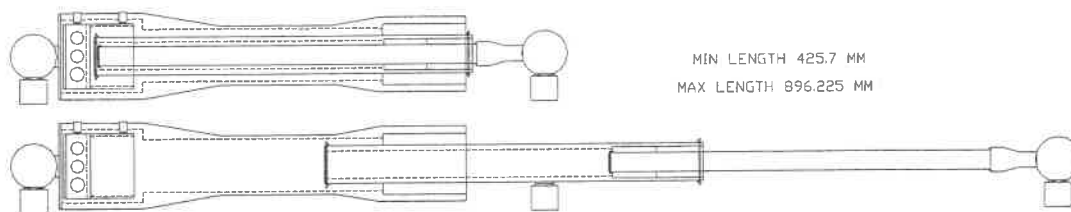


Figure 1: Schematic of laser ball bar used showing minimum and maximum extension positions.

The LBB is a multi-stage telescoping tube assembly with precision tooling balls located at the ends (Figure 1). A fiber-optically fed displacement measuring laser interferometer (DMI) is aligned with the motion axis of the LBB axis and measures changes in its length. However, in order to

measure distances between the ball centers, the LBB must be initialized to a known length. This is accomplished using the 3 step initialization process described in Figure 2.

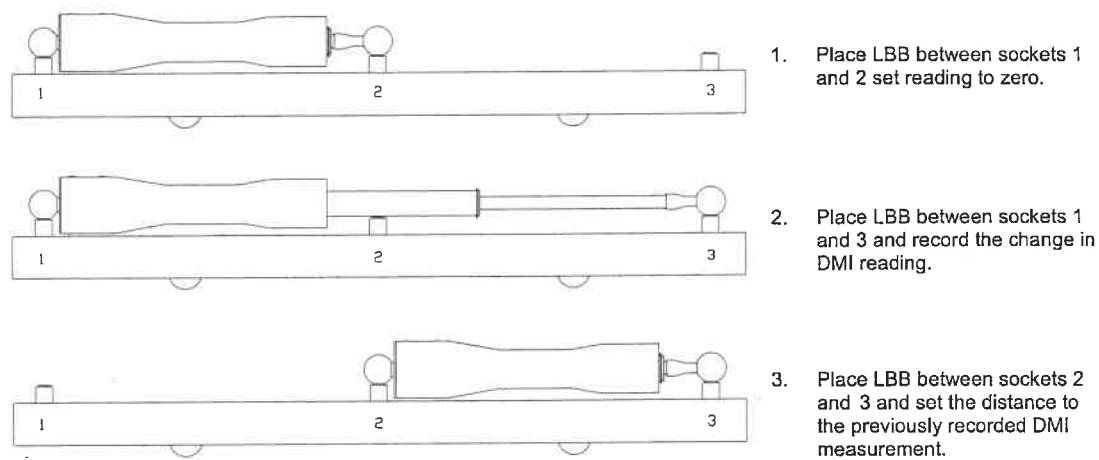


Figure 2: Three step process of initialization and calibration of the laser ball bar.

As shown in Figure 3, the LBB measures spatial coordinates using trilateration. Three magnetic sockets, designed to mate with precision tooling balls on the ends of the LBB, are fixed to the work table of the machine and another is held in the spindle in place of a cutting tool. These 4 sockets form the vertices of a tetrahedron. If the lengths of the six sides of the tetrahedron are measured, the coordinates of the "tool" socket relative to a coordinate system defined by the "base" sockets can be easily computed.

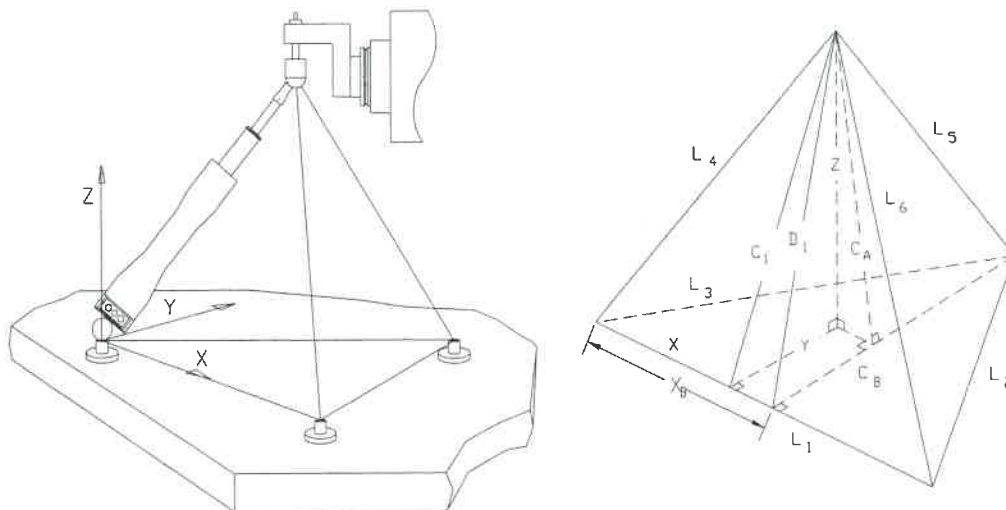


Figure 3: Diagram of coordinate system and setup for trilateration procedures.

2.1 LBB Measurand

The goal of the LBB measurement is to determine the spatial coordinates of the tool socket center via trilateration using LBB measurements of the lengths of the six sides of a tetrahedron. Since the DMI in the LBB can only measure length changes, the tetrahedron sides are obtained by adding the initialization length, L_0 , to the length change measured by the interferometer when the LBB is removed from the initialization fixture and placed between a pair of sockets, ΔL . Therefore, we must define multiple measurands in order to obtain an LBB measurement of the coordinates of the tool socket:

- a. ΔL : the true length change of the LBB from some initial configuration to a final configuration, as a function of the DMI reading,
- b. L_0 : the initialization length for a given initialization fixture design (function of ΔL),
- c. L_i : the lengths of the 6 sides of the tetrahedron (functions of L_0 and ΔL), and
- d. X_i, Y_i, Z_i : the coordinates of the target point (functions of L_i).

Thus, each measurement contains errors introduced during initialization, plus additional errors associated with changes in length of the LBB. These error sources are broadly categorized and modeled below. We estimate the uncertainty of LBB measurements using a Monte Carlo simulation of the measurement process. For each error source identified, we choose an appropriate statistical distribution based on the physical characteristics of the system, manufacturing tolerances, and engineering judgment.

3.0 Error Modeling

3.1 Mechanical errors

The LBB is composed of three rigid bodies which move relative to each other; 1) the base tube which holds the left ball and the DMI; 2) the center tube which carries the small tube; and 3) the small tube which holds the target retroreflector for the DMI and the right ball. Nominally, the two prismatic joints provide error-free, coaxial, straight-line motion, the interferometer axis is perfectly aligned with this motion, and there are no gravitationally or thermally-induced deformations of the bodies. In the real system, the joints are misaligned, do not provide perfect motions, and the individual bodies are subject to deformations.

The mechanical errors of the LBB are modeled using HTMs, treating the LBB as a two-axis mechanical system. Figure 4 shows a schematic of the LBB with coordinate systems assigned, and misalignments between components greatly exaggerated. CS1 is chosen to have its origin coincident with the left ball center, its X-axis parallel to the measurement axis of the DMI, and its Z-axis perpendicular to gravity. CS2 is attached to the center tube, with its origin at the center of the linear bearing which supports the small tube, its X-axis parallel to the bearing axis, and its Z-axis perpendicular to gravity. The origin of CS3 is at the right ball center with the X-axis passing through the apex of the retroreflector, and the Z-axis perpendicular to gravity.

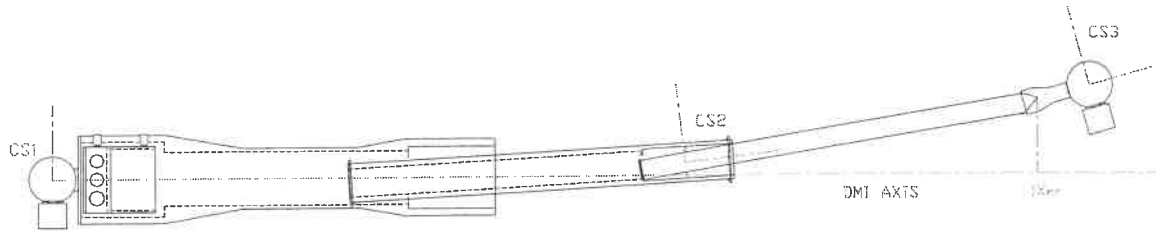


Figure 4: Schematic of the laser ball bar showing exaggerated misalignments and coordinate frames chosen for HTM computations.

The goal of the HTM model is to determine the actual coordinates of the retroreflector apex and the right ball center relative to CS1 for any configuration of the LBB. The X-coordinate of the retroreflector apex is the effective measurement point of the interferometer system, while the coordinates of the right ball center can be used to compute the actual LBB length. The nominal model is a function of the three section lengths measured along the DMI axis given by:

Equation 1

$${}^1T_{3,nom} = {}^1T_{2,nom} {}^2T_{3,nom} = \begin{bmatrix} 1 & 0 & 0 & L_a + L_b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & L_c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The mechanical and alignment errors are modeled using HTMs to give the actual relationship between CS3 and CS1.

Equation 2

$${}^1T_3 = {}^1T_2 E_2 {}^2T_3 E_3 = \begin{bmatrix} 1 & 0 & 0 & L_a + L_b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon_{z,2} & \varepsilon_{y,2} & \delta_{x,2} \\ \varepsilon_{z,2} & 1 & -\varepsilon_{x,2} & \delta_{y,2} \\ -\varepsilon_{y,2} & \varepsilon_{x,2} & 1 & \delta_{z,2} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & L_c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon_{z,3} & \varepsilon_{y,3} & \delta_{x,3} \\ \varepsilon_{z,3} & 1 & -\varepsilon_{x,3} & \delta_{y,3} \\ -\varepsilon_{y,3} & \varepsilon_{x,3} & 1 & \delta_{z,3} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3.1.1 Gravitational Deflections

For any given nominal pose of the ball bar, the gravitational deflections are obtained by modeling it as a simply supported, stepped beam. Castigliano's method is used to predict the displacement, δ_y , and Z-axis rotation, ε_z , of CS2 relative to CS1 and CS3 relative to CS2. The nominal dimensions and material properties of the LBB are used in the deflection calculations in combination with the nominal overall length and angle relative to gravity. In practice, the middle tube of the LBB is not fully constrained, and may take a range of positions whenever the total LBB length is between its minimum and maximum lengths. To account for this, the position of the middle tube is treated as an additional random input with uniform distribution bounded by the limiting positions.

3.1.2 Axis Motion Errors, Straightness and Rotation

The middle tube and small tube slide in linear ball bushings which are intended to produce pure rectilinear motion. However, manufacturing tolerances on the shafts cause deviations from this motion and result in translational errors in the Y and Z directions, δ_y and δ_z , and rotational errors about the Y and Z axes, ε_y and ε_z . The motion straightness errors (lateral error motions of the origins of CS2 and CS3 as the tubes extend) are estimated from the straightness tolerance for the middle tube and small tube prescribed in the design drawings, plus the lateral motion of the CS2 and CS3 origins due to the rotational errors in the bearing (normal distribution assumed). The values are perturbed within the ranges:

Middle tube:

$$\varepsilon_y = \pm 210 \mu\text{rad}$$

$$\varepsilon_z = \pm 210 \mu\text{rad}$$

$$\delta_y = \pm 5 \mu\text{m} + \varepsilon_z L_b$$

$$\delta_z = \pm 5 \mu\text{m} + \varepsilon_y L_b$$

Small tube:

$$\varepsilon_y = \pm 104 \mu\text{rad}$$

$$\varepsilon_z = \pm 104 \mu\text{rad}$$

$$\delta_y = \pm 2.5 \mu\text{m} + \varepsilon_z L_c$$

$$\delta_z = \pm 2.5 \mu\text{m} + \varepsilon_y L_c$$

3.1.3 Bearing Compliance

After construction, a slight misfit between the shaft and the linear bearing permits up to 0.35 mrad of rotation about the Z-axis for the LBB used in this study. The direction of this rotation is assumed to be governed by the gravity forces acting on the ball bar. When the ball bar is vertical, this angular error is assumed to be zero. This angular motion is added to the Z-rotation, ε_z , errors in the HTM, and the motion of the coordinate frame origin produced by this rotation is added to the Y direction displacement error, δ_y .

3.1.4 Cosine Errors

Initial misalignment between the DMI axis and the motion axis of the retroreflector causes a cosine error which is modeled as straightness error in the Y and Z directions proportional to the nominal displacement of the ball bar. The amount of cosine error is assumed to be normally distributed with zero mean and standard deviation of 262 μrad . The direction of misalignment relative to CS1 is treated as a uniformly distributed angle about the DMI axis with zero mean and range of $\pm\pi$ rad.

3.1.5 Retroreflector Coordinates in CS1

The coordinates of the retroreflector apex relative to CS1 can be found using the HTM model.

Equation 3

$${}^1\{P_{rr}\} = {}^1T_3^3\{P_{rr}\}$$

The X-coordinate of this vector is the projection of the retroreflector apex onto the DMI axis. Changes in this value are the displacements which will be measured by the interferometer. For a given DMI reading, we seek the true change in distance between ball centers, ΔL_{act} , as obtained from the HTM error model.

Equation 4-5

$$DMI = {}^1x_{rr} - {}^1x_{rr,0}$$

$$\Delta L_{act} = \sqrt{{}^1x_3^2 + {}^1y_3^2 + {}^1z_3^2} - \sqrt{{}^1x_{3,0}^2 + {}^1y_{3,0}^2 + {}^1z_{3,0}^2}$$

$$\text{where: } \begin{Bmatrix} x_3 \\ y_3 \\ z_3 \end{Bmatrix} = \begin{Bmatrix} {}^1T_3(1,4) \\ {}^1T_3(2,4) \\ {}^1T_3(3,4) \end{Bmatrix} \text{ are the coordinates of the origin of CS3 in CS1 during a measurement}$$

$$\text{and } \begin{Bmatrix} x_{3,0} \\ y_{3,0} \\ z_{3,0} \end{Bmatrix} = \begin{Bmatrix} {}^1T_{3,0}(1,4) \\ {}^1T_{3,0}(2,4) \\ {}^1T_{3,0}(3,4) \end{Bmatrix} \text{ are the coordinates of the origin of CS3 in CS1 at initialization}$$

3.2 Thermal deformations

The unsensed length between the left ball and the beamsplitter center is 22.685 mm. The unsensed length between the retroreflector apex and the right ball center is 34.350 mm. The coefficient of thermal expansion (CTE) for the stainless steel components making up these lengths was assumed to be 12 PPM/°C. Because the LBB is handled extensively during use, the temperature of this portion of the LBB is assumed to vary uniformly by up to $\pm 3^\circ\text{C}$ from the temperature at initialization. If the balls are assumed to be in sockets a fixed distance apart, thermal errors in the unsensed lengths will cause the DMI reading to change. Note that increase in temperature will increase the unsensed length and cause a decrease in the DMI reading. In the simulation, this error source is treated by perturbing the nominal coordinates of the retroreflector in CS3 by a normally distributed equivalent thermal error with zero mean and standard deviation of $0.62 \mu\text{m}$.

3.3 Interferometer errors

The interferometer system is subject to a number of well known error sources including: changes in the refractive index of air due to temperature, barometric pressure, humidity, and compositional fluctuations; imperfect optical elements, and polarization mixing. These errors are a function of the nominal optical path length of the measurement arm of the DMI, and are added to the mechanical and thermal displacement errors already described. Space limitations preclude detailed treatment of these errors here. The interested reader is referred to Schmitz, et al. [8,*9] and Bobroff [10]. For this simulation, the DMI errors are assumed to be normally distributed with zero mean and standard deviation of $0.125 \mu\text{m}$.

4.0 Measurand Uncertainty Computation

4.1 LBB Displacement Measurement, ΔL

When the LBB is displaced from some initial configuration to another configuration, the DMI reports a value for the length change. This is a "measurement result" for the ΔL measurand. The uncertainty of this result is "a parameter.....that characterizes the dispersion of values that could reasonably be attributed to the measurand" [1]. We estimate the true value of ΔL in the following manner:

- The initial distance between ball centers is assumed to be known, and the coordinates of the retroreflector apex in CS1 are calculated using the HTM model, while accounting for all of the mechanical error sources.

- b. When the LBB is moved to a new configuration, the nominal reported DMI value is corrected to account for the thermal and DMI errors.
- c. The corrected DMI value is used in conjunction with the HTM model of mechanical errors to obtain the new coordinates of the right ball center in CS1 which would give this DMI reading (i.e. the right ball position which would cause the retroreflector apex to be in the position corresponding to this DMI value).
- d. The right ball center coordinates are used to compute the new actual length of the LBB.
- e. The initial length of the LBB is subtracted from the actual length to give the true value of ΔL .

4.2 Initialization Errors

The first step in using the LBB is initialization to an absolute length. When the DMI system is turned on, it has no information about the distance from the beamsplitter to the retroreflector, or the distance between the ball centers. The initialization procedure provides an absolute distance between ball centers, to which subsequent displacements measured by the DMI are added. During initialization, the LBB itself is used to measure the distance, L_0 , between a pair of sockets on the initialization fixture. The LBB is then placed in these sockets and its length is set to this value.

There are two major sources of error in initialization. The LBB measurement of the socket to socket distance in the initialization fixture may be in error due to the displacement measurement errors of the LBB as described in Section 4.1. Also, the three sockets on the fixture may not be collinear (see Figure 5). For this simulation, the 3 sockets on the initialization fixture are assumed to be misaligned by up to $0.25 \mu\text{m}$. This error is assumed to be normally distributed with zero mean and standard deviation of $0.125 \mu\text{m}$.

For a given nominal distance between sockets 2 and 3 and the corresponding nominal DMI reading, the true distance between sockets 2 and 3, L_{23} , can be estimated as:

Equation 6

$$L_{23} = \left[\left(\sqrt{(L_{12} + \Delta L)^2 - dy^2} - L_{12} \right)^2 + dy^2 \right]^{1/2}$$

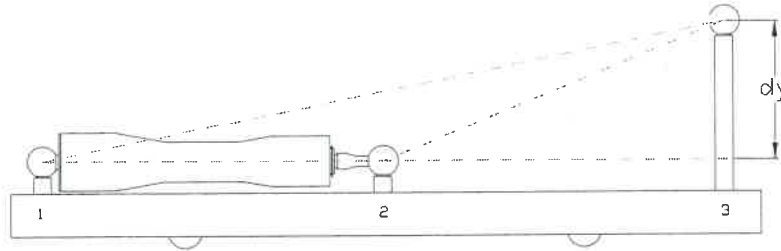


Figure 5: Exaggerated diagram of socket offset during initialization step.

4.3 Tetrahedron Side Measurement Errors

Following initialization, the LBB is placed between pairs of sockets which form the vertices of the measurement tetrahedron to measure the six inter-socket distances which are the tetrahedron edge lengths, L_i . Errors in these measurements are combinations of the initialization errors and displacement measuring errors described above.

$$L_i = L_0 + \Delta L_i \quad i = 1 \dots 6$$

Because the three base socket side lengths are typically measured once and then assumed to remain constant during all subsequent measurements, these lengths are assumed to undergo an additional thermal deformation during the measurement cycle. The thermal deformation is computed based on steel with a CTE = 11 PPM/°C, and a uniformly distributed $\pm 1^\circ\text{C}$ temperature variation.

4.4 Trilateration Errors

The spatial coordinates of the target point are computed from the tetrahedron edge lengths as outlined in Ziegert and Mize[4]. The statistical distributions of the side length measurements are used to compute the spatial coordinates.

Equation 7-9

$$\begin{aligned} x &= \frac{L_4^2 - L_5^2 + L_1^2}{2L_1} & c_b &= \sqrt{L_3^2 - x_b^2} \\ y &= \frac{d_1^2 - L_6^2 + c_b^2}{2c_b} & x_b &= \frac{L_3^2 - L_2^2 + L_1^2}{2L_1} \\ z &= \sqrt{c_1^2 - y^2} & d_1 &= \sqrt{c_1^2 + (x_b - x)^2} \\ & & c_1 &= \sqrt{L_4^2 - x^2} \end{aligned} \quad \text{where,}$$

5.0 Numerical Simulation of Uncertainty

The Monte Carlo simulation proceeds in the following steps:

- A nominal distance between initialization sockets 1 and 2 is assumed, and the HTM model is used to compute the coordinates of the retroreflector apex in CS1.
- A nominal DMI reading corresponding to the distance between sockets 2 and 3 is used to compute the coordinates of socket 3 in CS1 which would give this DMI reading.
- Equation 6 is used to determine the true distance between sockets 2 and 3, which is the true value of L_0 .
- Steps a...c are repeated 10,000 times and the mean and standard deviation of L_0 are computed.
- A set of nominal base socket locations is chosen along with the nominal coordinates of the target point to be measured. The nominal DMI readings corresponding to the six side measurements is computed, and used in the ΔL model to obtain the true LBB length change for each side. Each ΔL is added to value of L_0 selected from the distribution obtained in step d.
- Step e is repeated 10,000 times and the mean and standard deviations of the six side lengths are computed.
- Equations 7-9 are used to compute the coordinates of the target point from a set of side lengths chosen from the distributions obtained in step f. Repeat 10,000 times to obtain a point cloud of possible true values for the measurement point coordinates around the nominal point.
- Choose a new nominal point and repeat steps e through g.

Finally, after all of the data has been generated it is statistically analyzed to compute the variance in the measured coordinates at each point, and over the entire workspace.

6.0 Results and Conclusions

Sample results of a Monte Carlo simulation are shown for a tetrahedron with equilateral faces. The nominal DMI value chosen was equal to one half the maximum allowable extension of the LBB. Figure 6 shows the coordinates of the ideal case defined by these nominal lengths.

Figure 7a shows an enlarged image of the true location of the tool tip as evaluated using the error motions described above. Each data point represents one random configuration of

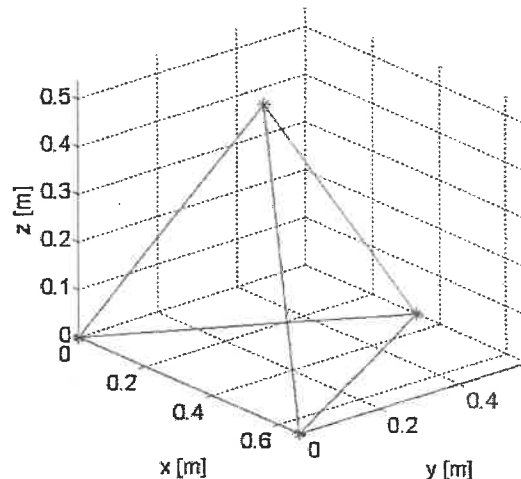


Figure 6: Nominal coordinates of the tetrahedron vertices used in the Monte Carlo simulation.

perturbed input values in the Monte Carlo program. These data points produce a point cloud which relates the possible true position of the tool tip to a given set of DMI readings. The cloud shown in Figure 7a is for a set of 100 data points, and the axes show the variation from the nominal value in each coordinate direction.

Figure 7b shows an ellipse that encompasses 66% of the data points for ten thousand trials. The radii in each coordinate direction are indicative of one standard deviation from the mean, $1.9 \mu\text{m}$ in x , $2.1 \mu\text{m}$ in y , and $1.0 \mu\text{m}$ in z .

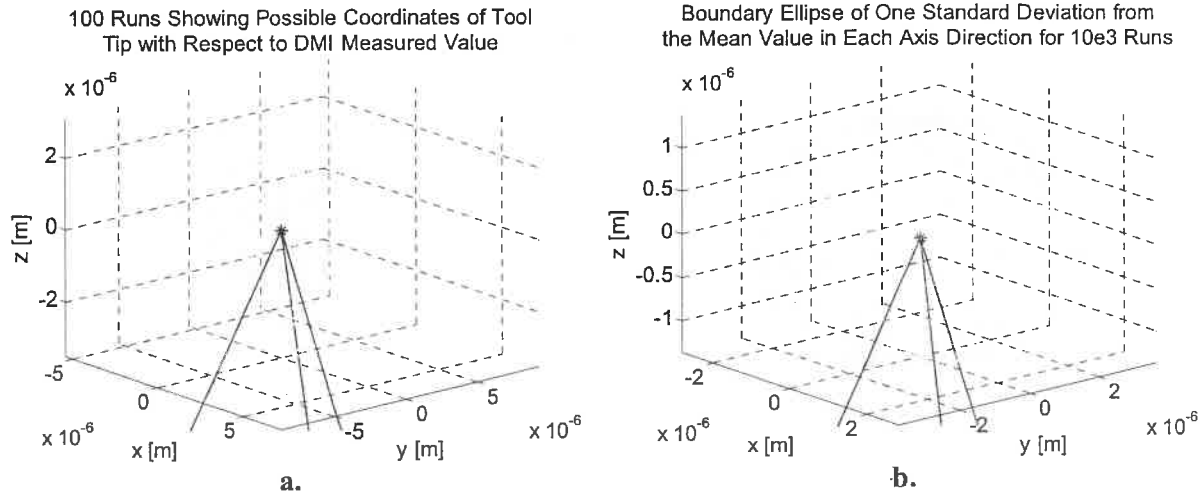


Figure 7(a, b): Uncertainty cloud shown around nominal position of tool tip as measured by DMI distances for a.) 100 trials and b.) one standard deviation of ten thousand trials.

With this study we have demonstrated a method for finding the uncertainty in 3-D coordinate measurements using the laser ball bar. The HTM method demonstrates how a complex, non-linear measurand can be simplified for analysis by decomposition of error motions which occur during measurement. Future research will explore how the uncertainty field varies within the work volume.

7.0 References

1. ANSI/NC SL Z540-2-1997, US Guide to the Expression of Uncertainty in Measurement, American National Standards Institute, 1997.
2. International Standards Organization, Guide to the Expression of Uncertainty in Measurement, 1995.
3. Davies, A. and Schmitz, T., 2003, Defining the Measurand in Radius of Curvature Measurements, Proceedings of the 18th ASPE Annual Meeting, October 26-31, Portland, OR (on CD).
4. Ziegert, J.C., and Mize, C.D., "The Laser Ball Bar: A New Instrument for Machine Tool Metrology", Precision Engineering: Journal of the American Society for Precision Engineering, v.16, no. 4, October 1994, pp. 259-267.
5. Srinivasa, N., Ziegert, J.C., and Mize, C.D., "Spindle Thermal Drift Measurement Using the Laser Ball Bar", Precision Engineering: Journal of the American Society for Precision Engineering, v. 18, no. 2/3, April, 1996, pp. 118-128.
6. Schmitz, T., Ziegert, J.C., "A New Sensor for the Micrometre-level Measurement of Three-dimensional, Dynamic Contours", Measurement Science and Technology, 10(2), February, 1999, pp. 51-62.
7. Schmitz, T. and Ziegert, J.C., Dynamic Evaluation of Spatial CNC Contouring Accuracy, Precision Engineering, 24/2, 2000, pp. 99-118.

8. Schmitz, T., and Beckwith, J., "An Investigation of Two Unexplored Periodic Error Sources in Differential-Path Interferometry", *Precision Engineering: Journal of the American Society for Precision Engineering*, v. 27, no. 3, July, 2003, pp. 311-322.
9. Schmitz, T., Evans, C., Davies, A., and Estler, W.T., 2002, "Displacement Uncertainty in Interferometric Radius Measurements", *Annals of the CIRP*, 51/1, 2002, pp. 451-454.
10. Bobroff, N., "Recent Advances in Displacement Measuring Interferometry", *Measurement Science and Technology*, v. 4, 1993, 907-926.

Acknowledgements

This work was supported in part by a grant from BWXT-Y12.

Methods for Improving the Accuracy and Repeatability in the Position of a Workpiece in Chucking

Jeongmin Byun, C.R. Liu
School of Industrial Engineering, Purdue University, West Lafayette, IN

Abstract

Significant improvements have been made in the performance of CNC machine tools due to the active research over the last several decades. However, the research and understanding on chucking accuracy have not been adequate. Nor has any systematic methods been reported for on chucking accuracy, especially in terms of the repeatability and the accuracy in the position of a chucked workpiece. The needs for the improvements in chucking accuracy has been existed since chucking has been the bottleneck in the form accuracy of a workpiece as the accuracy of machine tools improved significantly over the last several decades. However, the need was not very strong since turning has not been considered as the finishing process for the production of precision mechanical components such as bearings and shafts. After recent studies have demonstrated the benefits of hard turning in terms of surface integrity over other abrasive machining processes as a finishing process, a strong need emerged to improve chucking accuracy to enable the implementation of finish hard turning for the production of precision mechanical components. In this study, systematic methods for improving the accuracy and repeatability in the position of a chucked workpiece have been sought. Factors affecting the positioning accuracy and repeatability of a chucked workpiece have been identified. In addition, the characteristics of these factors, as well as their effect on chucking accuracy, were given. From the results, a chucking error map that summarizes the relations between these factors and the positioning error of a chucked workpiece was created. A series of error reduction experiments were carried out based on the results of the earlier works. The experimental results showed the possibility that the knowledge created in this study could be used effectively to improve the positioning accuracy and repeatability of parts in chucking and other work/tool holding operations.

Keywords: Chucking accuracy; Concentricity; Error budgeting; Error reduction; Finish hard turning

1. Introduction

There have been significant advances in the performance of CNC machine tools and the quality of cutting tools due to the active research over the last several decades. However, the research and understanding of chucking in term of the positioning accuracy and repeatability of a chucked workpiece has not been adequate. Nor has any systematic method been reported for improving chucking accuracy, especially in terms of the repeatability and the accuracy in the position of a chucked workpiece. The needs for the improvements in chucking accuracy has been existed since chucking has been the bottleneck in the form accuracy of a workpiece as the accuracy of machine tools improved significantly over the last several decades. However, the need was not very strong since turning has not been considered as the finishing process for the production of precision mechanical components such as bearings and shafts [1,2].

Hard turning is a term to describe the machining of hard materials, especially hardened steel. Traditionally, hard turning has not been considered as a finishing process for the production of these components due to the poor surface integrity and form accuracies. After recent studies have demonstrated the benefits of hard turning in terms of surface integrity, production efficiency and costs over other abrasive processes as a finishing process, a strong need has existed to improve chucking accuracy to enable the implementation of finish hard turning for the production of precision mechanical components [1,2,3].

In this work, systematic methods for improving the accuracy and repeatability in the position of a chucked workpiece have been sought. The goal of this paper is to develop an error map that shows the relations among the factors on affecting chucking error and the effects of these factors on chucking error. Another objective of this study is to develop methods for minimizing the chucking error of a cylindrical workpiece in order to help implement hard turning as a finishing process for the production of precision mechanical components.

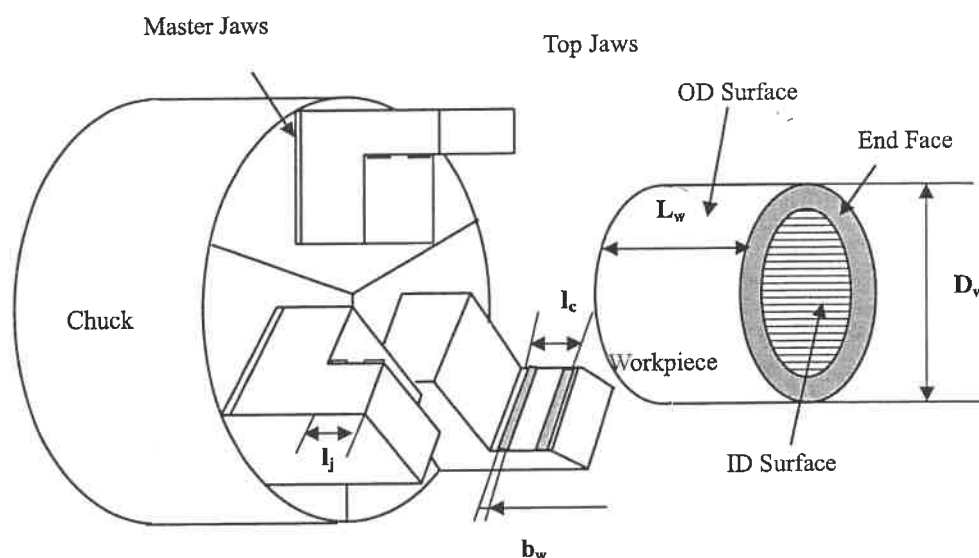
2. Error Budgeting

2.1 Basics of chucking error and research scope

Chucks are the devices that are most frequently used for the clamping of cylindrical workpieces. Their job is to center the workpiece to the spindle of a machine tool to the required degree of accuracy and to support it securely against the cutting force and to rotate it. If clamping is not sufficiently accurate, deviations in shape and position of the workpieces will occur and the dimensional and form accuracy of the machined workpiece will be degraded. This research covers the error budgeting and reduction of chucking error in a range of chucking/clamping situation where a cylindrical workpiece is overhung clamped. The workpiece is to be machined in turning that requires more than 1 set-up to finish. A cylindrical workpiece in this study includes a solid cylinder, a ring-shaped workpiece and a disk-shaped workpiece.

• Nomenclature

The majority of the following notations in Fig. 1 are based on the ANSI B5.8-1972, the American National Standard of Chucks and Chuck Jaws. Some of the notations that are not available directly from the Standard were adopted from other literatures on chucking [4,5,6,7].



D_w : Workpiece Diameter l_c : Clamping Length (Effective) L_w : Workpiece Length b_w : Width of Contact Band

Fig. 1. The definitions of parameters in chucking used in this work.

2.2 Definitions of chucking error

According to Pahlitzsch *et al.*[5], the deviations in shape and position which occur in the workpiece can be taken as a basis for defining chucking error. When a cylindrical workpiece is being machined, deviations from the ideal cylindrical shape may occur. When measured vertically to the axis, the deviation is called roundness error. When measured in the longitudinal section of the axis, the deviation is called straightness error (error of roundness and parallelism). Also, deviations in the position of unmachined surfaces can occur in the workpiece against the surfaces which develop during machining. The last deviation is called the deviation in the position of the workpiece.

Those deviations are measured in the form of radial run-out and axial run-out by means of experiments with clamped rotating test bars using probes. Even if the term, chucking error, may represent all sorts of errors caused by chucking operation in a broad sense, the chucking error of a workpiece conventionally represents the deviation in

the position of the workpiece due to inaccurate chucking [5]. According to Ema *et al.* [4], the chucking error was defined as the eccentricity of a test bar. In this paper, the positioning error of a workpiece in chucking is defined as the deviation of the center of unmachined surface from the center of the machined surface in the plane of measurement.

2.2 Identification of factors affecting the accuracy and repeatability of the position of a chucked workpiece

• Factors identified from literature review

A machine tool, a chucking device and cutting tools have been considered as separate entities, but their performance relies on the ability to function in harmony [7]. The work holding requires particular attention since its performance is critical. An improper chucking system can restrict the performance and accuracy of the machine tool and limit a machine's use. There have been a number of studies on chucking accuracy [4,5,6,7]. Table 1 summarizes the factors identified by these studies and their characteristics.

Table 1 Summary of factors affecting chucking error which were identified from the literature

Name of factor	Error type	Description of error
Radial displacement error of Individual jaws	Geometric	Internal wear of chuck components and backlash, geometric deviation of chuck components/clamp
Geometric error of workpiece	Geometric	Caused by the previous process ex) Diameter deviation, roundness error, squareness error
Taper in jaw alignment	Geometric /dynamic	Artificial taper : beneficial to chucking rigidity, Detrimental to chucking accuracy
Curvature difference between workpiece and jaw	Geometric	Higher curvature results to better jaw alignment and better roundness error
Self- alignment Of workpiece	Dynamic	Clamping force, moment arm, friction and angle of workpiece tilt as the factor on self-alignment
Selection of Locating surface	N/a	Clamping surface needs to be avoided For the clamping of workpiece with high l/d ratio, cylindrical surface should be considered
Non-symmetric deformation of jaw-workpiece	Geometric /dynamic	Caused by low rigidity of jaw-workpiece structure

• Factors identified by this study

The earlier studies [5,6,7] attributed the centering error of jaws to the geometric deviation of chuck components. In addition to the geometric errors, the authors have recognized that the centering error of a workpiece is caused by the factors described in Table 2.

Table 2 Summary of characteristics of factors affecting chucking error

Name of factor	Error type	Description of error
Jaw alignment error after jaw boring	Dynamic	Tilt or geometric deviation of boring plug causes the centering error in jaw alignment
Taper in jaw alignment	Dynamic	Higher clamping pressure with a chuck structure of lower rigidity can cause taper in jaw alignment.
Kinematic redundancy	Geometric	Mating of two imperfect parts (jaw and workpiece) can make the positioning of a workpiece non-repeatable.
Workpiece drop	Dynamic	Clamping of a long, heavy workpiece causes a workpiece drop in clamping.
Selection of locating surface	N/A	For the clamping of a long, heavy workpiece, a cylindrical surface needs to be used a locating surface.

2.3 Developing chucking error map

The factors discussed so far are summarized in Table 3, based on the stage of their occurrence in clamping. All the factors that occur in the same stage are listed in the same cell of the table.

Table 3 Factors affecting chucking error and their stages of influence

Stage of occurrence	Factors/errors affecting the eccentricity of chucked parts	Results
Preexisting Errors	▪ Deviation in shape of clamp transmission components ▪ Wear of clamp transmission component	Deviation in the radial displacement of individual jaws
Jaw boring	▪ Geometric error and tilt of jaw forming device ▪ Improper control of jaw boring pressure	Centering error and twist in the alignment of jaws after jaw boring Non-optimal tapering in jaw alignment
Workpiece mounting	▪ Kinematic redundancy, selection of locating face ▪ Workpiece Geometric Error, tapering in jaw alignment	Non-repeatable mating of workpiece and jaws Centering error of mounted workpiece
Jaw clamping	▪ Workpiece Drop ▪ Jaw-Workpiece Deformation, ▪ Self-Alignment of Workpiece	Centering error of chucked workpiece

A chucking error map was created, based on the results of error budgeting, as illustrated in Figure 3. To show the relation between these factors effectively, the error components that best represented the potential chucking error at each stage of clamping were defined as e_i .

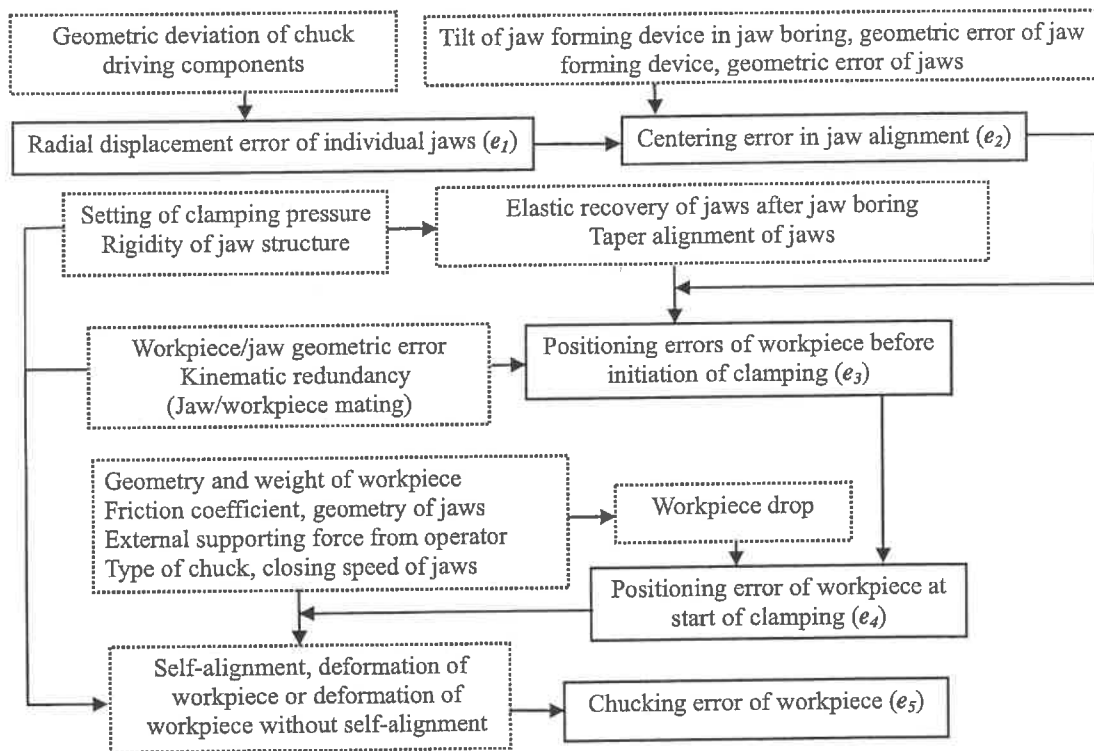


Fig. 2 Chucking error map

3 Developing Methods for Improving Chucking Accuracy

3.1 Methods for Minimizing Error Components in Chucking

• *Alignment error of jaws after jaw boring operation*

As shown in Table 3, two factors affect the positioning accuracy of the workpiece. The first one is the centering error in the alignment of jaws after jaw boring operation. The second is the excessive tapering in the alignment of jaws after jaw boring operation due to the inappropriate control of jaw clamping pressure [1]. Regarding the prevention of the tilt of the jaw forming device, an appropriate control of a clamping pressure is required to prevent the tapering of the surface on which the jaw forming device would be located. For example, a disk-shaped boring plug, which is used as a jaw forming device in most of the jaw boring operations, is located in a plug hole to withstand the clamping pressure during the boring operation of jaws. If the plug hole has a taper when it contacts to the 3 plug-hole surfaces, the disk would sit with a tilt since it will first contact with the bottom surfaces of a plug hole due to the inappropriate control of the clamping pressure. As an alternative, the use of a sphere-shaped disk plug can prevent the tilt of the plug by removing kinematic redundancy [1]. Another way to minimize the tapering in the jaw alignment is to raise the rigidity of the jaw structure by installing larger jaws if the situation permits. When lowering the clamping pressure in order to minimize the tapering in the alignment of jaws, clamping stability should be taken into consideration to prevent chatter in the machining.

• *Non-repeatability in positioning of workpiece due to kinematic redundancy*

In chucking, a workpiece that has a certain degree of geometric deviations is clamped by geometrically imperfect jaws with different matching orientations at each time of clamping. Under the conditions in which a cylindrical surface should be used as a major locator due to the higher length to diameter ratio of a workpiece, the final position of the workpiece becomes non-repeatable (kinematically redundant). Due to friction, the original position of the workpiece, which is far from optimal, would not be recovered in most cases once it is established at the time of mounting. In this case, an appropriate reduction of the contact area is required to make the positioning of the workpiece more repeatable and accurate. When a flat surface is used as a major locating surface, the non-repeatability in the positioning of a chucked workpiece can be minimized by arranging 3 point contact between the jaws and the workpiece unless there is no workpiece drop during clamping. However, the accuracy in the positioning of chucking is not improved significantly since it is mainly determined by other factors [1].

• *Chucking instability in form of workpiece drop*

In the chucking of a workpiece with a horizontal chuck spindle, gravity is not in favor of stabilizing the position of the chucked workpiece. According to the experiment, "workpiece drop" was observed in the clamping of a workpiece with 13.5 Kg weight and the ratio of L_w/D_w as 0.75. As the weight and the L_w/D_w increased, the instability increased to a point at which the support of a human operator cannot completely cover this instability. According to other works of the author [1], this drop increased as the weight and L_w/D_w increased. The l_c/L_w ratio should be high enough to prevent the drop of the workpiece in the clamping of this type of workpieces. However, a greater contact area increases the non-repeatability of the chucked workpiece. Therefore, efforts to minimize the kinematic redundancy should be made to improve the accuracy and repeatability of the chucked workpiece. As l_c/L_w increases, an appropriate attention should be paid on the change of actual major locating surface.

• *Positioning error of workpiece due to inappropriate selection of locating surface*

Two different types of surfaces can be used as a major locating surface in the clamping of a cylindrical surface, depending on the geometry and the weight of the workpiece. They are a flat surface (end face) or a cylindrical surface (OD/ID). If the length to diameter ratio of the workpiece is high, a cylindrical surface should be considered as a major locating surface to prevent the drop of the workpiece. Otherwise, a flat surface should be used to avoid all the problems of error propagation, as commented by Hocken *et al.*[6]. The more criteria and guidelines on the selection of a major locating surface are available in the other work of the authors. If there is no changeover, the use of a special chuck, which can support a pull-back operation, can be considered for the clamping of workpieces with a higher length to diameter ratio [1].

3.2 Experimental Verification

The methods based on the results of earlier works were tested using the outer rings of a taper-roller bearing. We have demonstrated that the methods could minimize the influence of these factors on the accuracy and the repeatability of the positioning of the chucked workpiece very effectively as shown in Figure 3a and 3b.

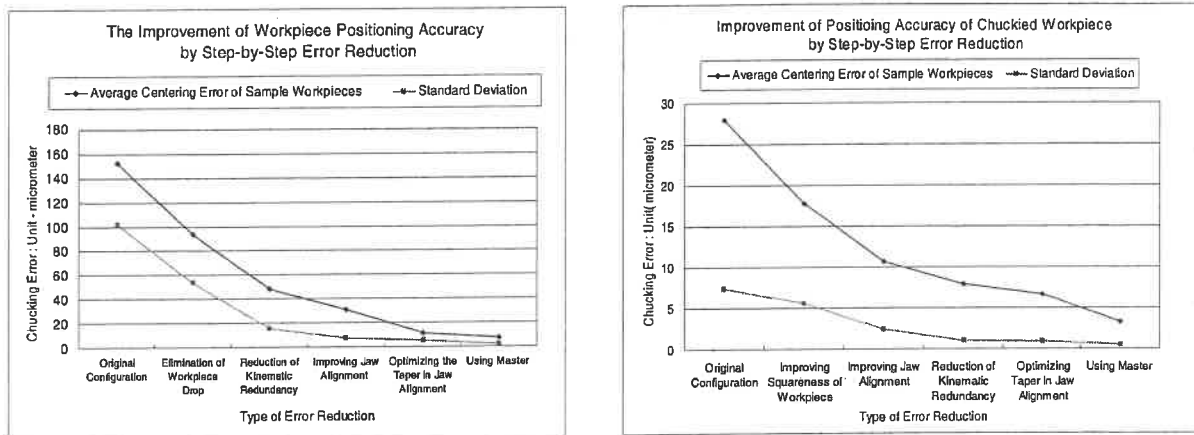


Fig.3a. The reduction results of chucking error by step-by-step approach for Workpiece Type I (OD surface as a major locating surface; length to diameter ratio 0.8, average weight 15.5 Kg).

Fig. 3b. The reduction results of chucking error by step-by-step approach for Workpiece Type II (End face as a major locating surface; length to diameter ratio 0.5, average weight 0.95 Kg).

4 Conclusions

In this work, factors that affect the chucking accuracy and repeatability of a cylindrical workpiece were identified. The characteristics of these factors and their effects on chucking errors were studied. From the result, a chucking error map was constructed. Methods for minimizing chucking error were developed, based on the earlier results. The methods were tested using the outer rings of a taper-roller bearing. We have demonstrated that the methods could minimize the influence of these factors on the accuracy and the repeatability of the positioning of the chucked workpiece very effectively. For the cases studied, we have achieved about 20 times improvement in chucking accuracy and 30 times in repeatability. The methods developed in this study can be applied to improve the accuracy and the repeatability of tools and parts.

List of Reference

- [1] J. Byun, Improving chucking accuracy of a cylindrical workpiece in finish hard turning, Ph.D. Dissertation, 2003, Purdue University.
- [2] Matsumoto, Y. (Timken Co); Hashimoto, F.; Lahoti, Surface integrity generated by precision hard turning, CIRP Annals - Manufacturing Technology, v 48, n 1, 1999, p 59-62
- [3] J.Y. Wang, C.R. Liu, Effect of tool flank wear on the heat transfer, thermal damage and cutting mechanics in finish hard turning. CIRP Annals - Manufacturing Technology 48 (1999) 53-58.
- [4] S. Ema, Chucking performance of a wedge-type power chuck, J. of Eng. for Ind. Trans. of ASME 116 (1) (1994) 70-77.
- [5] G. Pahlitzsch, W. Heddlwig, The clamping accuracy of three jaw chucks, Proc. The 8th Int. MTDR Conf., 1967, pp 97-118.
- [6] R.J. Hocken, Machine Tool Task Force, Technology of machine tools, Vol. 5, Machine Tool Accuracy, UCRL-52906-5 Lawrence Livermore Laboratory, UC Livermore, CA, 1980.
- [7] I. Ivashchenko, Some factors affecting the accuracy and rigidity of three-jawed chucks, Machines and Tooling 32 (1) (1961) 31-33.

CYLINDRICAL AND SPHERICAL COORDINATE TRANSFORMATIONS IN ERROR BUDGETS

Phil Smith

Mechanical Engineering Department, University of Kentucky, Lexington, KY *

R. Ryan Vallance

Department of Mechanical and Aerospace Engineering, The George Washington University,
Washington, DC †

Abstract

Error budgets are powerful tools for analyzing the uncertainty of positioning systems within machinery. A typical error budget employs homogeneous transformation matrices between Cartesian coordinate systems located along multiple pathways through the machine. Although Cartesian coordinates are appropriate for machines that employ linear motion, machines with rotary or spherical joints are better modeled with cylindrical or spherical coordinate systems. Employing Cartesian transformations in such cases may result in overstatement or understatement of the true error of the machine, particularly if second and higher order error terms are neglected. Instead, error terms must be presented as independent radial and angular errors, then converted to errors in a Cartesian coordinate system.

Keywords: error budget, error modeling, coordinate transformations, cylindrical coordinates, spherical coordinates

Introduction

The purposes of an error budget are to predict the uncertainty of positioning systems within machinery and to allocate error specifications to machine components. The error budget ensures that a machine design meets its specifications without being overbuilt. For modern machinery, where nanometer-level positioning uncertainties are expected, second-order effects must often be included within an error budget. Second order effects are also not negligible when one tolerance value is an order of magnitude or more larger than other tolerances in the same machine. Squared values of the larger tolerance may still be comparable in magnitude to the first order terms of the other tolerances.

A typical error budget combines a kinematic model of the machine with error estimates for machine components. The widely adopted approach summarized by Slocum [1] formulates a kinematic model using Cartesian coordinate systems and homogeneous transformation matrices for translations, rotations, or both. At the point where the "tool" contacts the "workpiece", a sensitive direction that critically affects cutting or measurement accuracy is defined, and insensitive directions are defined in directions orthogonal to the sensitive direction. Errors in the non-sensitive direction may have negligible effects and are often ignored.

Although the relationship between the sensitive and insensitive directions can be described by an orthogonal, Cartesian coordinate system, it is sometimes more accurate to express the sensitive and insensitive directions in cylindrical or spherical coordinates. Such is the case for applications with rotating sensitive directions like boring, where the sensitive direction coincides with the radial coordinate and the insensitive direction coincides with the angular coordinate. In such cases it is less appropriate to state that a particular Cartesian direction is insensitive, since errors are actually defined as deviations from a nominal radial value only.

This paper demonstrates that the approach of applying Cartesian coordinates and error values to each transformation may result in either an overstatement or understatement of error. As a consequence, machines may be overbuilt or emphasis may be applied to minimizing what amount to only secondary sources of error.

This paper introduces another approach where the connections between machine components dictate whether the use of Cartesian, cylindrical, or spherical coordinate systems is most appropriate. In this method, an important distinction is made between the mere rotation and transformation of Cartesian coordinate systems and the conversion from Cartesian to cylindrical or spherical coordinates. Errors for a cylindrical surface are evaluated in terms

* 151 RGAN Building, Lexington, KY 40506. psmith@engr.uky.edu.

† 738 Phillips Hall, 801 22nd St., N.W., Washington, DC 20052. vallance@gwu.edu.

of independent r , θ , and z values, rather than assuming independent error terms in x , y , and z . Similarly, errors in spherical coordinates are evaluated by examining independent r , θ , and ϕ error values.

The practice of always assuming a Cartesian coordinate system to represent sensitive and insensitive directions for a tool and workpiece can easily mask the interdependence of the errors, particularly if a Monte Carlo approach is used to determine the errors of the complete system. An example of the results from such a Monte Carlo simulation help illustrate the improvements offered by this new approach.

1. Cylindrical and Spherical Coordinate Transformations

Many machines contain connections or joints that are best modeled with cylindrical coordinates. Polishing or grinding wheels, cylindrical boring or turning machines, and many other designs have rotating elements constrained by spindles that establish axes-of-rotation. In the case where the axis-of-rotation is nominally aligned with the z axis, the transformations from cylindrical coordinates (r , θ , and z) to Cartesian coordinates (x , y , and z) are given by Eq. (1). These coordinates are defined in Fig. 1.

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \\ z &= z \end{aligned} \quad (1)$$

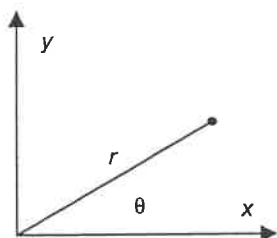


Fig. 1: Definition of polar coordinates.

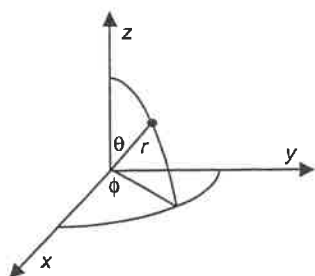


Fig. 2: Definition of spherical coordinates.

Some machines incorporate two rotational degrees of freedom located at the same point, indicating that spherical coordinates are the most suitable. The usual equations for conversion between r , θ , and ϕ are given by Eq. (2). The definitions of these angles are shown graphically in Fig. 2.

$$\begin{aligned} x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta \end{aligned} \quad (2)$$

2. Cylindrical Examples: Boring and Turning

Consider the production of concave and convex cylindrical surfaces produced by turning or boring as illustrated in Fig. 3 and Fig. 4, respectively. For turning operations such as Fig. 3, where the outside surface is being machined, it is apparent that errors in the angular position θ of the tool produce errors on the surface.

For Fig. 4, where the inside surface is being machined and the center of the tool is nominally coaxial with the axis-of-rotation of the workpiece, errors in the angular position θ of the tool should have no effect. Applying a Cartesian coordinate system to either geometry causes the distinction between these two cases to be obscured. In the coordinate system of the cylinder the sensitive direction is r , and the insensitive directions are θ and z . It is these variables, not x , y , and z , that are the truly independent variables indicating the position of the tool.

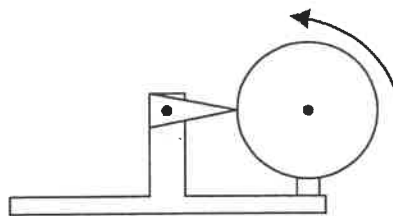


Fig. 3: Turning operation with a fixed tool and rotating workpiece.

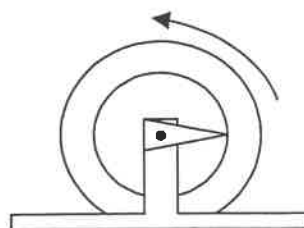


Fig. 4: Boring operation with a fixed tool and rotating workpiece.

For Fig. 4, the tool holder may be designed to provide a highly accurate r value, with less accuracy in its θ positioning. If the conversion from r and θ values to x and y values include second order terms, with x as the sensitive direction, errors in θ may appear to adversely affect the sensitive direction x . This is not a problem with the design but rather with the use of Cartesian coordinates for sensitive and insensitive directions.

With many machines, errors in the radial direction Δr and angular direction $\Delta\theta$ may be independent; encoder resolution, for instance, should be represented as a contributor to $\Delta\theta$ and independent of Δr . Depending on the order of the error model, error in the x direction (Δx) and y direction (Δy) may be statistically correlated even though the errors $\Delta\theta$ and Δr are independent. This correlation, and whether the system error is overstated or understated, depends on the geometry of the machine as shown in subsequent sections.

3. First Order Cylindrical Example

The simplest example to demonstrate this correlation is to apply first order coordinate transformations to the tool in Fig. 4. We assume a finite radius r , a nominal value of $\theta=0$, and an angular error of $\Delta\theta$. Errors in the alignment of the two axes-of-rotation, and any errors in the out-of-plane (z) direction are ignored. First order, small angle approximations are given by Eq. (3).

$$\begin{aligned}\sin \Delta\theta &\approx \Delta\theta \\ \cos \Delta\theta &\approx 1\end{aligned}\quad (3)$$

For this first order approximation the rotation error matrix is shown in Eq. (4), and the translation matrix is shown in Eq. (5). These two matrices are combined to give the x and y coordinates as shown in Eq. (6). Second order terms are dropped, to maintain consistency within this approach.

$$T_{1-ord} = \begin{bmatrix} 1 & \Delta\theta & 0 & 0 \\ \Delta\theta & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$T_{trans} = \begin{bmatrix} 1 & 0 & 0 & r + \Delta r \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$T_{1-ord}T_{trans} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} r + \Delta r \\ r\Delta\theta \\ 0 \\ 1 \end{bmatrix} \quad (6)$$

From these results it is apparent that the errors Δx and Δy are related to the original, independent variables Δr and $\Delta\theta$, as shown in Eq. (7). These two equations are written in matrix form to highlight the fact that, since the errors Δx and Δy are orthogonal combinations of Δr and $r\Delta\theta$, they are linearly independent. From a purely linear algebra standpoint, they are orthogonal because the dot product of the two row vectors of the matrix in Eq. (7) is zero.

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta r \\ r\Delta\theta \end{bmatrix} \quad (7)$$

4. Second Order Cylindrical Example

The next step to check for correlation is to apply second order approximations to the coordinate transformations of Fig. 4. We once again assume a finite radius r , a nominal value of $\theta=0$, and an angular error of $\Delta\theta$. The second order small angle approximations are given by Eq. (8).

$$\begin{aligned}\sin \Delta\theta &\approx \Delta\theta \\ \cos \Delta\theta &\approx 1 - \Delta\theta^2/2\end{aligned}\quad (8)$$

Using these second order approximations, the rotation error matrix is shown in Eq. (9). The translation matrix is unchanged from the previous example, and so Eq. (5) can be used. These two matrices are combined to give the x and y coordinates shown in Eq. (10). Third order and higher terms are dropped in this case.

$$T_{2-ord} = \begin{bmatrix} 1 - \Delta\theta^2/2 & \Delta\theta & 0 & 0 \\ \Delta\theta & 1 - \Delta\theta^2/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$T_{2-ord}T_{trans} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} r + \Delta r - r\Delta\theta^2/2 \\ r\Delta\theta + \Delta r\Delta\theta \\ 0 \\ 1 \end{bmatrix} \quad (10)$$

From these results it is apparent that the errors Δx and Δy are related to the original, independent variables Δr and $\Delta \theta$, as shown in Eq. (11). Now the errors Δx and Δy are nonlinear combinations of Δr and $\Delta \theta$, and are no longer independent.

$$\begin{aligned}\Delta x &= \Delta r - r\Delta\theta^2/2 \\ \Delta y &= r\Delta\theta + \Delta r\Delta\theta\end{aligned}\quad (11)$$

To show the correlation between Δx and Δy graphically, three different sets of random errors are generated for the tool shown in Fig. 3 or 4. For this data the radius r of the tool is set to 40 cm, the error Δr is ± 0.01 cm, the angle θ is 0, and the angular error $\Delta\theta$ is ± 0.05 radians.

Physical measurement of the errors Δx and Δy would reveal a range of -0.06 to $+0.01$ cm in x , and ± 2 cm in y . Figure 5 shows a uniform distribution of Δx and Δy within these ranges. Each point plotted represents a possible location of the toolpoint relative to the ideal position of $\{0, 0\}$.

An improved method of modeling the errors is to measure Δr and $\Delta\theta$, then use the first order approximation from Eq. (7) to convert these errors to Δx and Δy coordinates. Figure 6 shows uniformly distributed Δr and $\Delta\theta$ variables. Since the first order approximation maintains the independence of the variables, the distribution is still uncorrelated but in a much smaller range.

The third and final method of modeling the errors is to utilize the second order approximations for Δx and Δy , found in Eq. (11). With the same ranges on Δr and $\Delta\theta$, Fig. 7 reveals how the correlation affects the randomly distributed original error variables.

A key question in this study is whether the omission of the second order terms results in an overstatement of error or an understatement of error. Unfortunately, the direction of the effect depends on the unique geometry of the problem to be analyzed. The two machines in Fig. 3 and Fig. 4 are contrasting examples of this effect.

For the convex cylinder of Fig. 3, the data points in Fig. 7 fall further from the ideal surface than the data points in Fig. 6. Neglecting the second order terms for this machine would therefore result in an understatement of error. For Fig. 4, the points in Fig. 7 more closely approximate the interior of the cylinder, therefore the omission of the second order terms results in an overstatement of error. In both cases, the assumption of independent Δx and Δy values as shown in Fig. 5 results in the largest overstatement of error, and should never be applied.

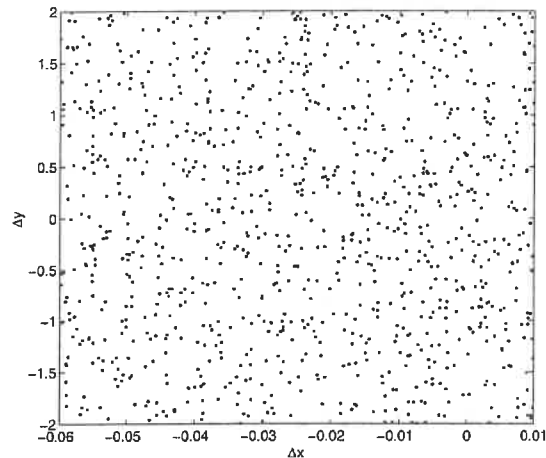


Fig. 5: Example of uncorrelated Δx and Δy .

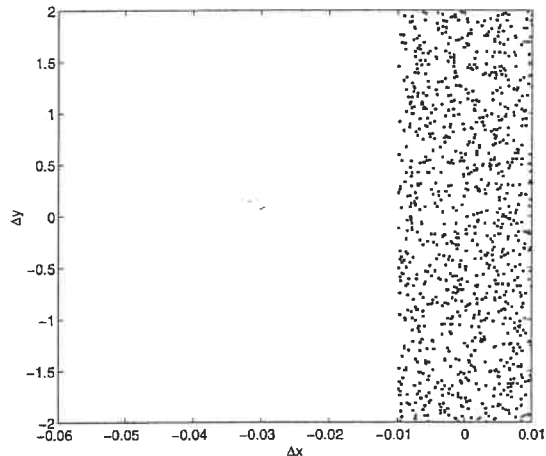


Fig. 6: Example of uncorrelated Δx and Δy , generated from first order approx. for Δr and $\Delta\theta$.

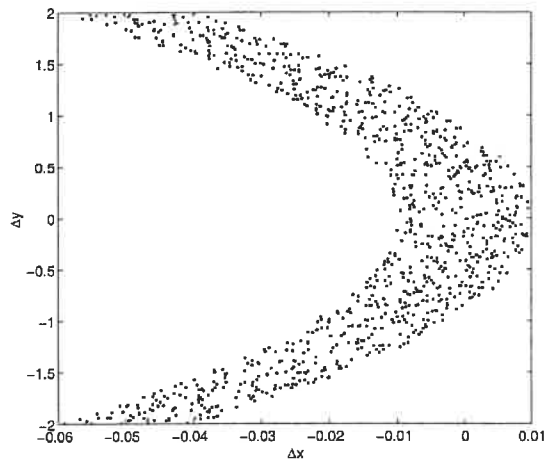


Fig. 7: Example of correlated Δx and Δy , generated from second order approximations for Δr and $\Delta\theta$.

5. Generalized First Order Cylindrical Results

The previous example was somewhat specific in that the machine tool was positioned horizontally, i.e. the tool was not rotated through a finite angle θ . To determine the first and second order equations for Δx and Δy more generally, the transformation in Eq. (12) must be applied. The first order approximations for these angles are shown in Eq. (13).

$$T_{rot} = \begin{bmatrix} \cos(\theta + \Delta\theta) & -\sin(\theta + \Delta\theta) & 0 & 0 \\ \sin(\theta + \Delta\theta) & \cos(\theta + \Delta\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$$\begin{aligned} \sin(\theta + \Delta\theta) &\approx \sin \theta + \Delta\theta \cos \theta \\ \cos(\theta + \Delta\theta) &\approx \cos \theta - \Delta\theta \sin \theta \end{aligned} \quad (13)$$

Multiplying this rotation matrix with the translation matrix of Eq. (5) produces a result that can be written as shown in Eq. (14).

$$T_{rot} T_{trans} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x + \Delta x \\ y + \Delta y \\ 0 \\ 1 \end{bmatrix} \quad (14)$$

The x , Δx , y , and Δy results can be written once again in matrix form, as shown in Eq. (15). Since the dot product of the two rows of the matrix in Eq. (15) is zero, the vectors Δx and Δy are shown to be orthogonal.

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} \Delta r \\ r\Delta\theta \end{bmatrix} \quad (15)$$

6. Generalized Second Order Cylindrical Results

The second order, general solution is more complex. The cosine and sine approximations, to second order, are shown in Eq. (16). The resulting equations for Δx and Δy are shown in Eqs. (17) and (18). These are neither linear nor orthogonal.

$$\begin{aligned} \sin(\theta + \Delta\theta) &\approx \sin \theta + \Delta\theta \cos \theta - \frac{\Delta\theta^2}{2} \sin \theta \\ \cos(\theta + \Delta\theta) &\approx \cos \theta - \Delta\theta \sin \theta - \frac{\Delta\theta^2}{2} \cos \theta \end{aligned} \quad (16)$$

$$\begin{aligned} \Delta x &= \Delta r \cos \theta - r\Delta\theta \sin \theta \\ &\quad - \Delta r\Delta\theta \sin \theta - \frac{r\Delta\theta^2}{2} \cos \theta \end{aligned} \quad (17)$$

$$\begin{aligned} \Delta y &= \Delta r \sin \theta + r\Delta\theta \cos \theta \\ &\quad + \Delta r\Delta\theta \cos \theta - \frac{r\Delta\theta^2}{2} \sin \theta \end{aligned} \quad (18)$$

7. Generalized First Order Spherical Results

The same technique applied to two-dimensional cylindrical transformations can be applied to the general case of spherical coordinates as defined in Fig. 2. The rotations to convert from cylindrical coordinates r , θ , and ϕ , to Cartesian coordinates x , y , and z , are shown in Eqs. (19) through (21). This example requires two rotations and a translation, to be multiplied as shown in Eq. (22).

$$T_1 = \begin{bmatrix} \cos(\phi + \Delta\phi) & -\sin(\phi + \Delta\phi) & 0 & 0 \\ \sin(\phi + \Delta\phi) & \cos(\phi + \Delta\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (19)$$

$$T_2 = \begin{bmatrix} \cos(\theta + \Delta\theta) & 0 & -\sin(\theta + \Delta\theta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\theta + \Delta\theta) & 0 & \cos(\theta + \Delta\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (20)$$

$$T_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & r + \Delta r \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (21)$$

$$T_1 T_2 T_3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x + \Delta x \\ y + \Delta y \\ z + \Delta z \\ 1 \end{bmatrix} \quad (22)$$

For the first order approximation of sine and cosine, once again Eq. (13) is used. Second order and higher terms are discarded, with the resulting error terms Δx , Δy , and Δz as shown in Eq. (23).

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \theta \sin \phi \\ \sin \theta \sin \phi & \cos \theta \sin \phi & \sin \theta \cos \phi \\ \cos \theta & -\sin \theta & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta r \\ r\Delta\theta \\ r\Delta\phi \end{bmatrix} \quad (23)$$

As before, the rows of the matrix in Eq. (23) can be considered as three separate vectors. Unlike the cylindrical example, however, these vectors are not orthogonal to one another. In particular the dot product of Δx and Δy is nonzero, indicating a correlation between these two variables.

This correlation indicates that, for a machine with independent errors in Δr , $\Delta\theta$, and $\Delta\phi$, the related Cartesian errors Δx , Δy , and Δz will always be interdependent. Whether this interdependence results in an understatement or an overstatement of error again depends on the particular geometry of the machine.

8. Generalized Second Order Spherical Results

As in the case of cylindrical transformations, the most accurate calculation of the Cartesian errors Δx , Δy , and Δz must include second order error terms in Δr , $\Delta\theta$, and $\Delta\phi$. The rotation matrices of Eqs. (19) through (21) are again applied, but the second order approximations for sine and cosine are taken from Eq. (16). Terms higher than second order are dropped, and the resulting equations for the Cartesian errors are shown in Eqs. (24) through (26).

$$\begin{aligned} \Delta x = & \left(\Delta r - \frac{r}{2} \Delta\phi^2 - \frac{r}{2} \Delta\theta^2 \right) \sin\theta \cos\phi + \\ & (r\Delta\theta\Delta\phi) \cos\theta \sin\phi - (r\Delta\phi + \Delta r\Delta\phi) \sin\theta \sin\phi \\ & + (r\Delta\theta + \Delta r\Delta\theta) \cos\theta \cos\phi \end{aligned} \quad (24)$$

$$\begin{aligned} \Delta y = & \left(\Delta r - \frac{r}{2} \Delta\phi^2 - \frac{r}{2} \Delta\theta^2 \right) \sin\theta \sin\phi + \\ & (r\Delta\theta\Delta\phi) \cos\theta \cos\phi + (r\Delta\phi + \Delta r\Delta\phi) \sin\theta \cos\phi \\ & + (r\Delta\theta + \Delta r\Delta\theta) \cos\theta \sin\phi \end{aligned} \quad (25)$$

$$\Delta z = \left(\Delta r - \frac{r}{2} \Delta\theta^2 \right) \cos\theta - (r\Delta\theta + \Delta r\Delta\theta) \sin\theta \quad (26)$$

Although these equations are admittedly convoluted, terms are grouped in an attempt to gauge the relative effects of different terms. If the radius r and the angular errors are large, second order terms such as $r\Delta\theta\Delta\phi$ may have a significant effect. Conversely, second order terms that are deemed negligible can be selectively eliminated from Eq. (24) through (26).

10. References

- [1] Slocum, A. H., *Precision Machine Design*, Society of Manufactu

9. C

We ha
comp
spher
to Car
evalu
result

For th
single
appro
the Ca
appro
betwe
was a
result
mach
the m

For th
conve
and C
of the
betwe
indica
indep
be mo
in the
calcul

It is
Fig. 4
coinc
In a c
will a
transf
the m
errors
neces

It ha
autho
probl
exam
effect
positi
no sin
matri
Gener
may b

As before, the rows of the matrix in Eq. (23) can be considered as three separate vectors. Unlike the cylindrical example, however, these vectors are not orthogonal to one another. In particular the dot product of Δx and Δy is nonzero, indicating a correlation between these two variables.

This correlation indicates that, for a machine with independent errors in Δr , $\Delta\theta$, and $\Delta\phi$, the related Cartesian errors Δx , Δy , and Δz will always be interdependent. Whether this interdependence results in an understatement or an overstatement of error again depends on the particular geometry of the machine.

8. Generalized Second Order Spherical Results

We have shown that the errors for many machines components are best described using cylindrical or spherical coordinates. As these errors are converted to Cartesian coordinates, care must be taken in evaluating the dependence or independence of the resulting Cartesian errors.

For the example of cylindrical machining with a single point tool, it was shown that the first order approximation resulted in independent variables in the Cartesian coordinate system. The second order approximation, however, showed a strong correlation between the errors in the Cartesian coordinates. It was also shown that overlooking this correlation may result in an overstatement or understatement of the machine error, depending on the exact geometry of the machine.

For the general case of spherical coordinates, the conversion equations between spherical errors terms and Cartesian error terms are shown. Even in the case of the first order equations, correlation is found between some of the Cartesian variables. This indicates that the errors for a machine with independent spherical degrees of freedom should first be modeled as spherical coordinate errors. The errors in the Cartesian coordinate system must then be calculated from the spherical errors.

It is assumed throughout the cylindrical analysis of Fig. 4 that the axis of rotation of the tool was coincident with the axis of rotation of the workpiece. In a complete design analysis, errors in positioning will affect this alignment, and additional transformation matrices are necessary. Depending on the magnitude of the translations and the rotation errors, second order terms may or may not be necessary throughout the model.

It has also been brought to the attention of the authors that it is often necessary to solve the reverse problem of the transformations discussed here. For example, it is often necessary to determine what effect errors in Δx , Δy , and Δz will have on positioning in spherical coordinates. Unfortunately, no simple transformation matrices exist to produce a matrix of Δr , $\Delta\theta$, and $\Delta\phi$ values from Δx , Δy , and Δz . General equations to study the effect of these errors may be considered at a future time.

10. References

- [1] Slocum, A. H., *Precision Machine Design*, Society of Manufacturing Engineers, Dearborn, MI (1992).

As in the case of cylindrical transformations, the most accurate calculation of the Cartesian errors Δx , Δy , and Δz must include second order error terms in Δr , $\Delta\theta$, and $\Delta\phi$. The rotation matrices of Eqs. (19) through (21) are again applied, but the second order approximations for sine and cosine are taken from Eq. (16). Terms higher than second order are dropped, and the resulting equations for the Cartesian errors are shown in Eqs. (24) through (26).

$$\Delta x = \left(\Delta r - \frac{r}{2} \Delta\phi^2 - \frac{r}{2} \Delta\theta^2 \right) \sin \theta \cos \phi + (r \Delta\theta \Delta\phi) \cos \theta \sin \phi - (r \Delta\phi) \sin \theta \sin \phi + (r \Delta\theta + \Delta r \Delta\theta) \cos \theta \cos \phi \quad (24)$$

$$\Delta y = \left(\Delta r - \frac{r}{2} \Delta\phi^2 - \frac{r}{2} \Delta\theta^2 \right) \sin \theta \sin \phi + (r \Delta\theta \Delta\phi) \cos \theta \cos \phi + (r \Delta\phi) \sin \theta \cos \phi + (r \Delta\theta + \Delta r \Delta\theta) \cos \theta \sin \phi \quad (25)$$

$$\Delta z = \left(\Delta r - \frac{r}{2} \Delta\theta^2 \right) \cos \theta - (r \Delta\theta) \sin \theta + (r \Delta\phi) \sin \theta \cos \phi \quad (26)$$

Although these equations are admittedly convoluted, terms are grouped in an attempt to gauge the relative effects of different terms. If the radius r and the angular errors are large, second order terms such as $r \Delta\theta \Delta\phi$ may have a significant effect. Conversely, second order terms that are deemed negligible can be selectively eliminated from Eq. (24) through (26).

Evaluation of the geometrical uncertainty of helix deviation measurements

using the Monte Carlo simulation

Hideyuki Tanaka¹, Makoto Shimajo², Takashi Sato³

¹ NMI/AIST, ² Univ. of Electro-Communications, ³ OSAKA SEIMITSU KIKAI Co., Ltd.

Introduction

In the evaluation of uncertainty, the usual approach is to write down a model equation and apply the law of propagation of uncertainty to it. However, when the measurand is a complex derived quantity, the model equation occasionally involves an inverse function that cannot be solved analytically. It is not possible to directly apply the law of propagation of uncertainty to such a situation.

In this study, the uncertainty of a quantity involving an inverse function that cannot be solved analytically is evaluated by using the Monte Carlo simulation. Prior to this, we constructed a program that calculates the uncertainties of the profile deviation of gear measurement¹⁾. We pick up the helix deviation measurement of the gear in this study.

Principle and model equation of helix deviation measurement

The principle of helix deviation measurement is shown in Fig. 1.

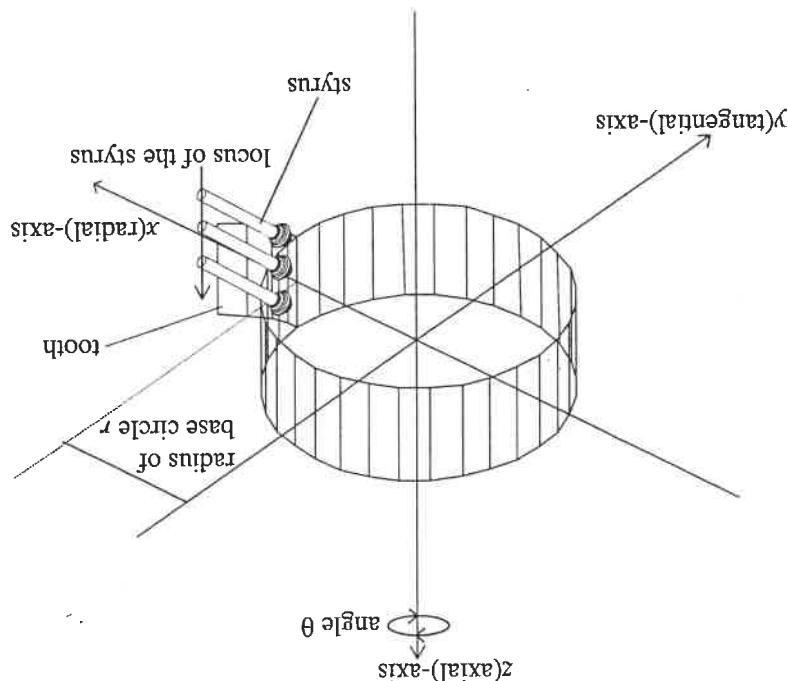


Fig. 1: Principle of helix deviation measurement

When the helix deviation of the gear is measured, the y-position of the stylus may be obtained. Subsequently, we calculate the model equation of the y-position of the stylus.

Position of the gear tooth:

The ideal position of the gear tooth (x_0, y_0, z_0) is expressed by the following equations:

Here, P^{x_rand} and P^{z_rand} are the measurement errors of the x- and z-direction, respectively. P^{yaw} represents the yawing of the z-axis. Accordingly, the x- and z-position of the stylus can be calculated.

$$\begin{aligned} X_2 &= X_1 + P^{x_rand} + P^{z_yaw} \\ Z_2 &= Z_1 + P^{z_rand} \end{aligned}$$

Let the x-position and z-position of the stylus that includes random errors be X_2 and Z_2 , respectively.

Here, X^{slopeV} is the inclination of the inner gage against the reference plane, and X^{slopeH} is the inclination of the linear gage for the reference plane. The other constants obey this law.

$$\begin{aligned} X_1 &= X_0 \cos(X^{slopeV}) \cos(X^{slopeH}) - X^{slopeH} + Z_0 \sin(Z^{slopeX}) \\ Z_1 &= Z_0 \cos(Z^{slopeX}) \cos(Z^{slopeY}) \end{aligned}$$

Let the x-position and z-position of the stylus that includes the geometrical errors be X_1 and Z_1 , respectively.

P^x is the fixed measurement point. Identically, P^x is a static value; and P^z is the displacement in the direction of measurement; therefore, the value of P^z varies.

$$\begin{aligned} X_0 &= P^x \\ Z_0 &= P^z \end{aligned}$$

Let the identical x-position and z-position of the stylus be X_0 and Z_0 , respectively.

Position of the stylus:

Here, $rot(\phi, \theta_1)$ represents the rotation matrix, and θ_1 is the sum of the identical value of rotation and the error of rotation. (x_2, y_2, z_2) represents the actual position of the gear tooth.

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = rot(\phi, \theta_1) \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$$

The spindle of the gear inclines and the gear rotates. Let the inclination of the gear spindle be ϕ , the angle of rotation be θ_1 , and the position of the gear tooth after inclination and rotation be (x_2, y_2, z_2) .

$$\begin{aligned} x_1 &= x_0 + x^{ecc} \\ y_1 &= y_0 + y^{ecc} \\ z_1 &= z_0 \end{aligned}$$

The Gear measurement machine has an eccentricity and a origin setting error. Let the sum of those errors be (x^{ecc}, y^{ecc}) , and the position of the gear tooth that includes the eccentricity and origin setting error be (x_1, y_1, z_1) . Here, r is the radius of the base circle, a is the parameter of the involute function, and W_0 is the width of the gear tooth.

$$\begin{aligned} x_0 &= r(\cos a + a \sin a) \\ y_0 &= r(\sin a - a \cos a) \\ -W_0 &\leq z_0 \leq W_0 \end{aligned}$$

About the simulation and results

The factors of the measurement uncertainty taken into consideration included: eccentricity of the rotation axis, origin setting error, slope of the rotation axis, measurement uncertainty of the rotary encoder, measurement uncertainty of the scales, pitching of the z (axial)-axes, and slopes of x (radial)-, y (tangential)- and z (axial)-axes.

The preceding individual uncertainties were measured at the University of Electro-Communications and Osaka Seimitsu Kikai Co., Ltd²⁾. The results of the measurements are listed in Table 1, and the simulation conditions are listed in Table 2.

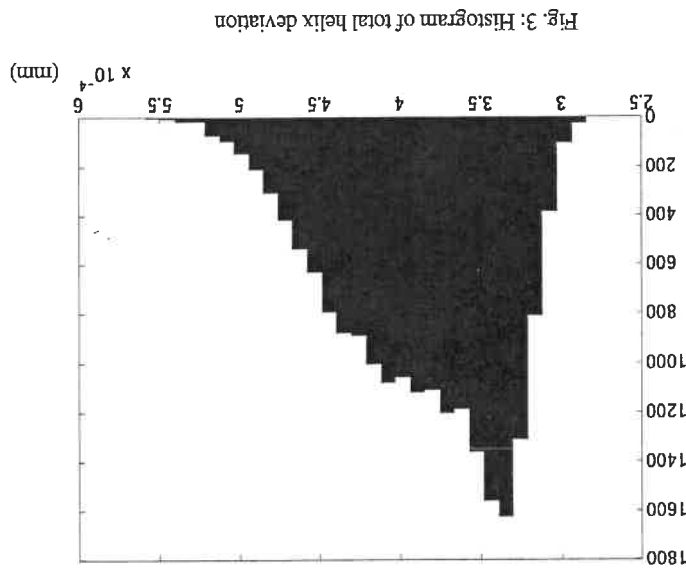
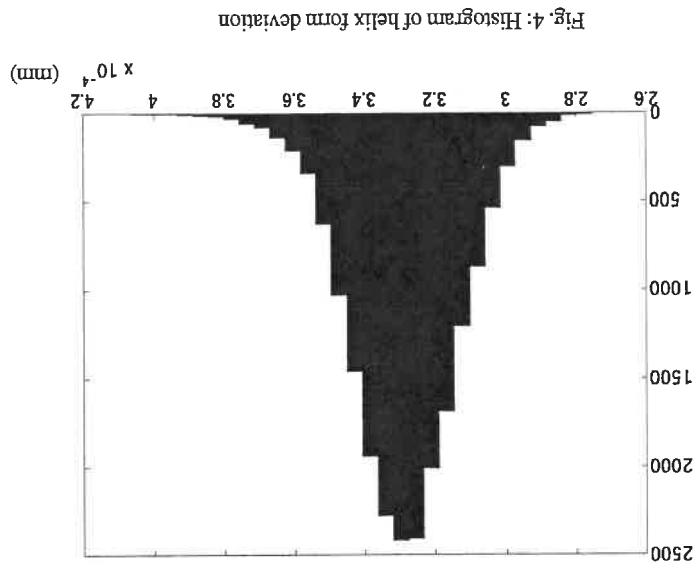
The individual uncertainties listed in Table 1 are limited values except for the length measurement uncertainty (y-axis direction). The distributions of individual uncertainties, except for the length measurement uncertainty (y-axis direction), are assumed to be rectangular, and the distribution of the length measurement uncertainty (y-axis direction) is assumed to be normal.

Table 1: Individual uncertainties

Source of uncertainties	Value
Eccentricity (x-axis direction)	0.51 μm
Eccentricity (y-axis direction)	0.51 μm
Origin setting error (x-axis direction)	0.32 μm
Origin setting error (y-axis direction)	0.12 μm
Length measurement uncertainty (x-axis direction)	0.3 μm
Length measurement uncertainty (y-axis direction)	0.0296 μm
Length measurement uncertainty (z-axis direction)	0.3 μm
Pitching of z-axis	0.15 μm
Angle of x-axis for the reference plane	0.001 $\mu\text{m}/100 \text{ mm}$
Angle of x-axis against the reference plane	0.001 $\mu\text{m}/100 \text{ mm}$
Angle of y-axis for the reference plane	0.000245 rad
Angle of y-axis against the reference plane	0.000118 rad
Angle of z-axis for x-axis	0.0005 $\mu\text{m}/150 \text{ mm}$
Angle of z-axis for y-axis	0.0005 $\mu\text{m}/150 \text{ mm}$
Uncertainty of an angle of rotation	0.00000242 rad
Slope of rotation axel (x-axis direction)	0.0000060 rad
Slope of rotation axel (y-axis direction)	0.0000154 rad

Table 2: Simulation conditions

Number of partitions of measurement diameter	100 points/10 mm
Radius of base circle	49.4 mm
Model of tooth	300,000 points per line, 0.1 μm
Repetitions	20,000



The helix deviations evaluated by us correspond to the total helix deviation, the helix form deviation, and the helix slope deviation, which were defined by ISO 1328-1: 1995 "Cylindrical gears—ISO system of accuracy—Part 1: Definitions and allowable values of deviations relevant to corresponding flanks of gear teeth." The histogram of total helix deviation is shown in Fig. 3, the histogram of helix form deviation is shown in Fig. 4, the histogram of helix slope deviation is shown in Fig. 5, and one of the results which was obtained from 20,000 simulations is shown in Fig. 6. Consequently, the uncertainty of the total helix deviation measurement was $0.47 \mu\text{m}$, the uncertainty of the helix form deviation measurement was $0.35 \mu\text{m}$, and the uncertainty of the helix slope deviation measurement was $0.22 \mu\text{m}$, where the uncertainties are expressed in terms of expanded uncertainties corresponding to the 95% level of confidence.

gear measurement machine. Report of WG2 of committee for gear measurement accuracy (2002) (in Japanese).

2) Committee for gear measurement accuracy. Performance evaluation and estimation uncertainty of a high precision Society for Precision Engineering, pp. 395-398 (2003)

ing instruments by using the Monte Carlo simulation. Proceedings of the Eighteenth Annual Meeting of the American Yuzaki, Shogo Kato, Yoshiharu Shibuya, and Takayuki Kondo. Evaluation of measurement uncertainty in gear measur-

1) Hideyuki Tanaka, Makoto Shimajo, Makoto Kajitani, Takashi Sato, Ryuhei Takeda, Masasaki Kojima, Masatoshi

Bibliography

We thank the concerned member of the committee for gear measurement accuracy for investigating gear measurement traceability and calibrations, and for helpful discussion.

Acknowledgment

This study estimates the expanded uncertainty of the measurement of the gear helix profile whose model equations include the inverse involute function. However, this model does not include whole uncertainties; this point must be improved upon. In future, we will structure a simulation for helical gear and pitch deviations.

Conclusions

Fig. 6: Example of result of simulation

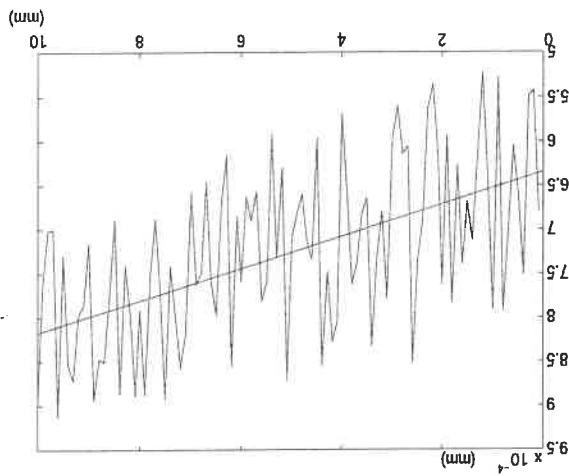
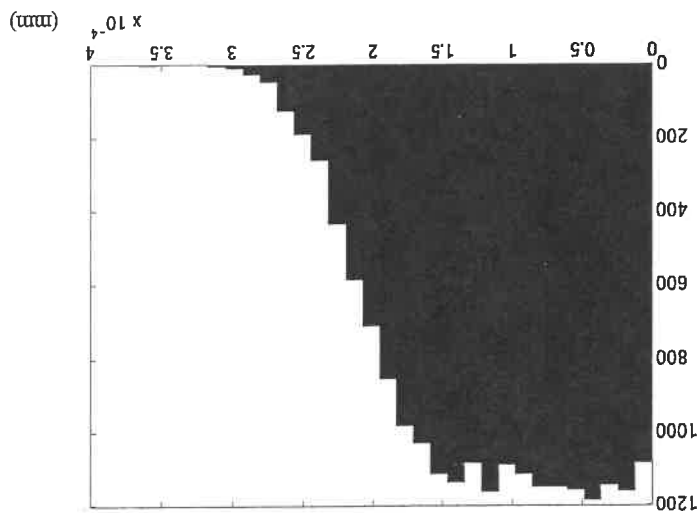


Fig. 5: Histogram of helix slope deviation



The "Virtual CMM" a software tool for uncertainty evaluation – practical application in an accredited calibration lab

Michael Trenk*, Matthias Franke**, Dr. Heinrich Schwenke**

*FEINMESS GmbH & Co. KG

**Physikalisch-Technische Bundesanstalt,

Department Coordinate Metrology, 38116 Braunschweig, Germany

Contact:

Key Words: coordinate measuring machines (CMMs), CMM calibration, dimensional artifacts

Abstract

As a practicable method to evaluate measurement uncertainties for almost any measurement task on CMM PTB has developed the "VIRTUAL CMM" concept (VCMM). This method is based on Monte-Carlo simulations of the error behavior of a real CMM. Meanwhile the VCMM concept has been integrated in the software of two German CMM manufacturers and within the German Calibration Service (DKD) four laboratories are accredited to calibrate workpieces using this technique. FEINMESS as one of the accredited labs is presenting its practical experiences in this paper.

Introduction

A complete measurement result should consist of the measured value and its uncertainty. The statement of a reliable task specific uncertainty in coordinate metrology is very challenging. This is due to the characteristic of a CMM as a multi-purpose measuring instrument, the large number of uncertainty contributors and their complex kind of propagation. Measurement tasks that include a number of different geometrical features and datums can generally not be described by classical uncertainty budgets. A way to solve this problem is to use Monte-Carlo based simulation tools to calculate task specific uncertainties as described in [3] or [5]. Together with project partners from academia and industry PTB has integrated the simulation method in commercial software: The "Virtual CMM". Today, accredited calibration labs in Germany use this technique to perform calibrations of prismatic parts as an essential link in the traceability chain down to the shop floor.

The role of the Virtual CMM in the traceability chain

One of the main tasks of PTB as the National Metrology Institute of Germany is to ensure the comparability and reliability of measurement results by making measurements traceable to the SI units. In the field of 3D metrology for prismatic parts, the traceability can be achieved by a calibration chain as outlined in Fig. 1. Starting from the meter definition and its realization by wavelength standards, it propagates traceability to industrial measurements through a chain set up by 2D-artifacts, calibrated CMMs and calibrated workpieces. The most complex chain link is the CMM, which has to take the step from simple 2D artifacts to complex workpieces, which can incorporate a number of complex features with all their task specific contributors. The Virtual CMM simulation software can generate valid uncertainties for all of the measured or derived features and therefore the process of measurement and simulation can be regarded as a traceable calibration. After calibration, these workpieces can be used for experimental uncertainty determination according to ISO 15530 [1]. In order to supply industry with calibrated workpieces PTB established a calibration service under the aegis of the DKD (German Calibration Service) in cooperation with competent calibration laboratories. These laboratories are now accredited according to ISO 17025 to calibrate arbitrary prismatic parts as reference workpieces, traceable in the strict sense. As a result, industry now can be supplied with calibrated workpieces in order to establish traceability of measuring equipment on the shop floor level.

CMM / Environment	• systematic deviations of the slide-ways • uncertainties due to calibration • thermal deformation of the slideways • thermal expansion of the scales • drift effects
Probing process	• direction-dependent behavior • uncertainties of the stylus • calibration • uncertainties when using a probing system with several styli
Workpiece	• thermal expansion of the workpiece • roughness of workpiece surfaces

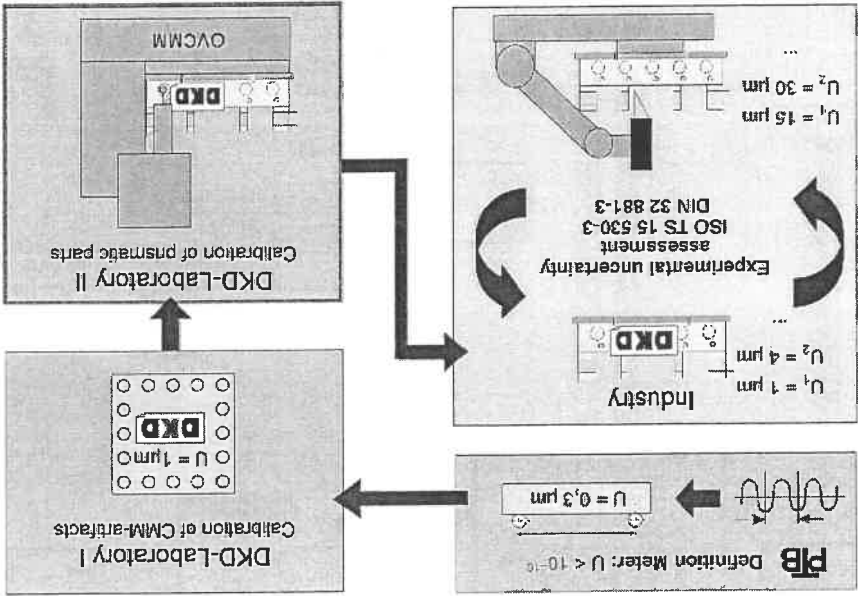
Essential for the performance of the VCM is the provision of input parameters to describe all relevant contributions to the error behavior of the CMM. Currently the following contributors are considered: measurement report of the CMM software.

with a coverage of 95%. The actual measured value and the related uncertainty are stated together in the results for each measurand. This interval is interpreted as the extended measurement uncertainty U ($k=2$) samples, a simple statistical routine is calculating an interval that includes approximately 95% of all re-200) of coordinates and obtains a representative sample of potential measurement results. From these The evaluation software of the CMM calculates results for these additional sets (typically around 100- and random deviations which are added to the nominal coordinates.

corded. From the recorded set of points further sets are generated by the VCM by generating systematic rates. The evaluation software of the CMM calculates the actual result from the set of coordinates re- of the VCM. The real CMM probes the workpiece at specified points and records a set of point coordi- packages (CALYPSO by Zeiss and QUINTOS by Messstechnik Weizlar). Fig. 2 illustrates the operation means of a "virtual experiment". The CMM simulator is now integrated in two German CMM software budget, the base of the uncertainty evaluation. However the problem is not solved analytically, but by measuring process and the uncertainty sources affecting it are, analogous to the classical uncertainty niques are well established in many domains of nature and engineering sciences. The model of the Basis of the VCM method is the emulation of the measuring process by statistic simulation. These tech-

Basic principle of the Virtual CMM (VCM)

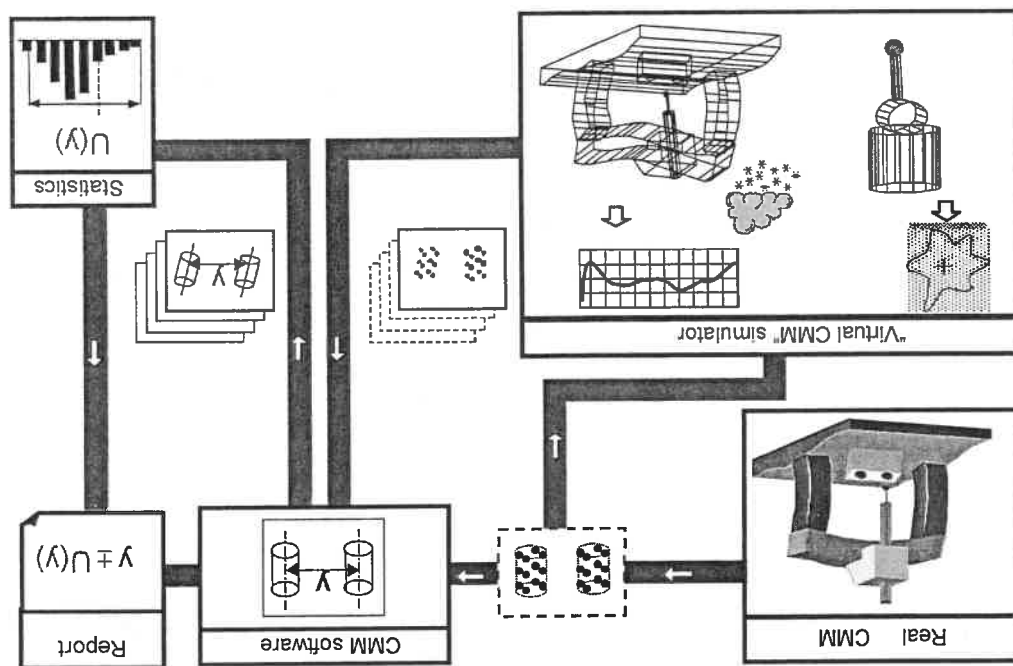
Fig. 1: Calibration chain for CMM



The input parameters for the simulation software must be determined in advance by specially designed calibration procedures and long term observations. In the following the basic input parameters and the operations for their determination are briefly described. To support the creation of input parameter files PTB has developed the software VCMTool. It can be understood as the user's interface for the parameterization of the error model of the VCM. VCMTool serves e.g. for entering geometrical parameters of the CMM, temperatures and gradients, material constants and additional constraints. It is also possible to check visually how certain parametric errors are modeled, e.g. changing straightness due to thermal gradients. To assess the systematic parametric error functions a method based on measurements of calibrated ball or hole plates is used [2]. For this purpose, PTB provides the analysis software called KALKOM. For a full error map it is necessary to measure the plate in at least six positions in the CMM's volume. An upcoming alternative method is the application of high accuracy tracing interferometers (developed by PTB and NPL). This new procedure is less time consuming and also suitable for large scale CMM [7]. To determine error contributions caused by changes of the ambient conditions a long term monitoring of temperature and gradient is essential. This is done using multi-channel temperature logging systems with at least four sensors in the surrounding of the CMM. After the qualification of a stylus with the standard routine in the CMM control software, there usually still are residual systematic and random deviations which have an effect on the measurement uncertainty. The relating model actually used in the VCM is developed for mechanically contacting (tactile) probing systems. It describes the behavior in a general form by harmonic functions. This general model has to be individually scaled for each employed stylus. The appropriate parameters are assessed by an additional measurement on the calibration sphere. For this purpose, QUINTOS provides a procedure for probing points according to a specified pattern and transmitting results for storage in the VCM database.

Determining input parameters

Fig. 2: Data flow of the integrated VCM (simplified)



To increase the variety of the measuring tasks additionally, also position and orientation of the test cylinder in the measuring volume, the number and distribution of measuring points and stylus configuration have been varied for the test. In Fig. 4 measurement results are displayed as deviations from the calibration values as well as the uncertainties determined by simulation (error bars) and the uncertainties of the calibration values (boxes) exemplarily for one of the participating laboratories. Altogether the results show that the uncertainty ranges of all measured values under inclusion of the calibration values overlap. A measure for the compliance can be the F_N value [9], which sets measuring deviations and the uncertainties into relationship. Consistence can be assumed if the condition $|F_N| \leq 1$ is true. For the accreditation this has to be the case for at least 95% of all measured features.

Fig. 4: Deviations of measurement results from calibrated values (both with uncertainties)

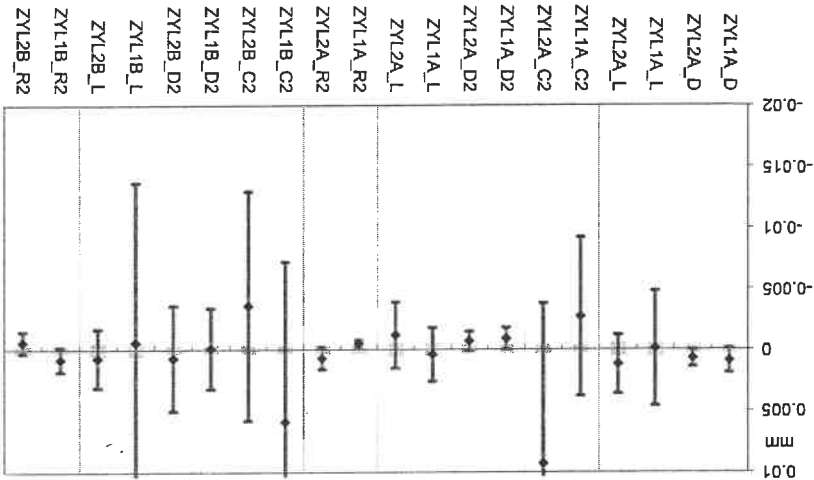
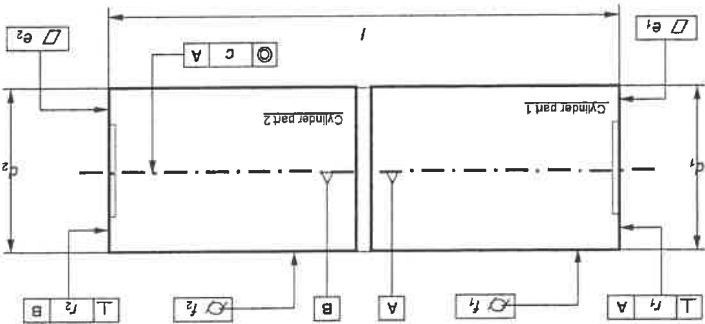


Fig. 3 : Various measurement tasks embodied on a cylinder



For proving the proper installation, parameterization and application of VCM by the calibration labs user comparative measurements on calibrated artifacts have been performed as one prerequisite for becoming accredited. Mainly test cylinders have been used because they can embody a set of different measurement tasks (diameters, length, squaredness, coaxiality of different areas of the cylinder, see Fig. 3) and all this geometrical features can be calibrated with sufficient accuracy on CMM using reversal techniques [8].

Concept verification

Interim check and re-verification

For the application in a calibration lab it has to be guaranteed that the assumptions, made for creating the VCM input parameters still correspond to the actual conditions. This is realized by regular monitoring measurements of calibrated artifacts, in order to verify the entire system as "black box". The measurement task has to be quick and simple but at the same time sensitive to all important uncertainty influences, such as changes of scale factor due to temperature, changes of squareness due to changing temperature gradients or collisions, de-adjustment of the probing system. Two procedures, complying these requirements, are the measurement of ball plates in two crossed positions (according VDI 2617 sheets 5) and the measurement of the so-called ball cube. The decision, whether the VCM still calculates the "correct" uncertainties, can be determined by comparing the values and uncertainties of the actual test with the calibrated values. If there is no contradiction the input parameters can be assumed still as valid. In a research project with CMM manufacturers, metrology service providers and users the PTB develops at present an internet-based system, which makes a "remote controlled" monitoring possible by accredited labs and so in particular release small and medium enterprises of this task [6].

Practical examples of calibration tasks

Within its the first year as a DKD accredited laboratory FEINMESS performed about 25 calibrations of prismatic parts for customers with uncertainties calculated by "Virtual CMM". Fig. 5 shows some examples to illustrate the variety of calibration tasks. Beside the DKD calibrations, VCM at FEINMESS is also used regularly for optimizing part programs regarding probing strategy and datum setting.

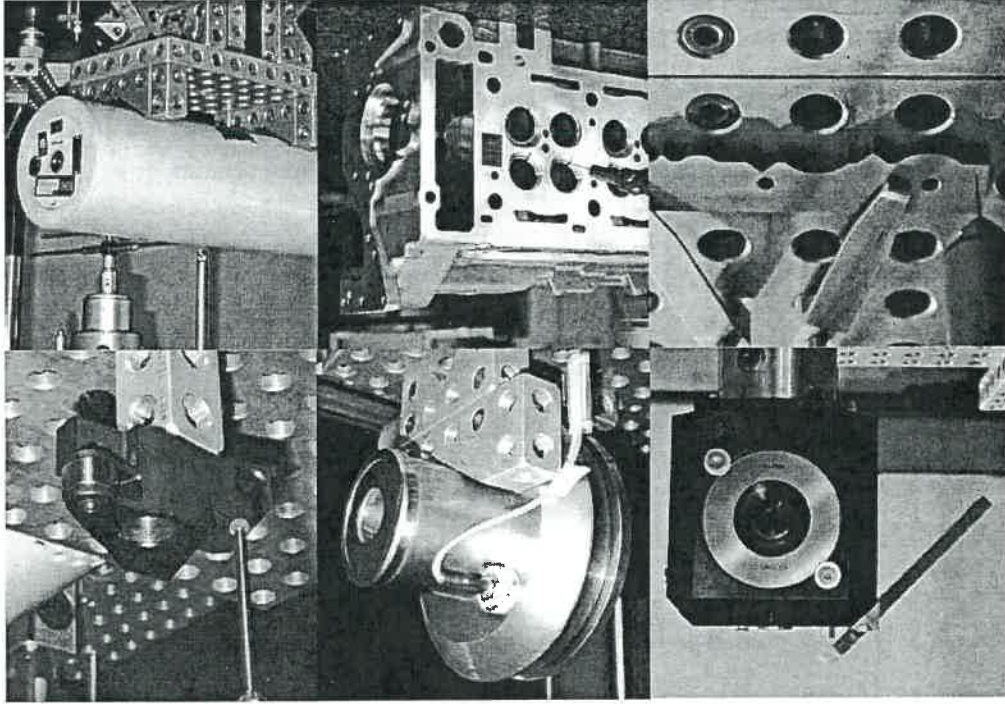


Fig. 5: Examples for calibration tasks

Conclusion and outlook

The "Virtual CMM" concept is a milestone on the way to achieve traceability in the field of coordinate metrology. This tool is meanwhile integrated and commercially available in two CMM software packages and there is a growing interest from other vendors. Further development targets the extension of the functionality, e.g. integration of models for scanning and rotary tables and also the transfer of the method to other types of measuring devices such as contour tracers. Another important issue is the simplification of the procedures, in particular for defining the input parameters, in order to attract a wider range of users beside the calibration labs.

References

1. ISO/TS 15530-3:2004 Geometrical Product Specifications (GPS) -- Coordinate measuring machines (CMM): Technique for determining the uncertainty of measurement -- Part 3: Use of calibrated workpieces or standards
2. *Trapet, E.; Waldele, F.*: A reference object based method to determine the parametric error components of coordinate measuring machines and machine tools. In: Measurement 9 (1991), S. 17-22
3. Trapet, E., Waldele, F.: The Virtual CMM Concept. Advanced Mathematical Tools in Metrology, World Scientific Publ. Comp., 1996, p 238 – 247
4. Trapet, E., Franke, M., Härtig, F., Schwenke, H., Waldele, F., Cox, M., Forbes, A., Delbressine, F., Schellekens, P., Trenk, M., Meyer, H., Moritz, G., Guth, Th., Wanner, N.: Traceability of Coordinate Measurements According to the Method of the Virtual Measuring Machine, Final Project Report MATI-CT94-0076, PTB-report F-35, Part 1 and 2, ISBN 3-89701-330-4, 1999
5. Phillips, S. D., et al.: The Calculation of CMM Measurement Uncertainty via the Method of Simulation by Constraints, Precision Engineering, 1997a. 16: p. 443-446.
6. Kreis Dr., R.; Franke, M.; Schwenke Dr., H.; Waldele Dr., F.: Improving Traceability of Complex Measuring Systems in Industry, IDW2002 - International Dimensional Workshop, Knoxville (USA), 06.-09.05
7. Schwenke, H., Franke, M., Hannaford, J., Schneider, C.-T.: Calibration of CMMs and machine tools by parametric multilateration, IMEKO 8th International Symposium on Measurement and Quality Control Erlangen, GERMANY, 2004, October 11-15 (to be published)
8. Osawa, S., Busch, K., Franke, M., Schwenke, H.: Multiple orientation technique for the calibration of cylindrical workpieces on CMMs, Precision Engineering (to be published soon)
9. WECC Interlaboratory comparisons, WECC document 15, 1992

A Framework for Uncertainty Evaluation of GPS Standard-chain

J.X. Wang, L.M. Ma, X.Q. Jiang, Z.G. Xu, Z. Li

Dept. of Mechanical Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China

Abstract The uncertainty is used as an economic tool to enable optimum allocation of resource amongst specification, manufacturing and verification. For a given GPS standard-chain, the key question is to calculate the compliance uncertainty. The determination of compliance uncertainty is a complex question because specification uncertainty and measurement uncertainty arise from many causes and propagate through the GPS standard-chain. A framework for compliance uncertainty of GPS standard-chain is proposed in this article, which is based on the modeling of GPS standard-chain. According to ISO 17450-2, a GPS process should be either in default state or in special state. The biggest difference between the two states is that whether the specification operator is accordant with the verification operator. Aiming at the two states, the flow for the computation of compliance uncertainty is given respectively. It enables us to generate compliance uncertainty on the verification of GPS standard-chain, which makes the acceptance or refusal of feature characteristic can be conducted in a quantitative way, so the veracity of verification could be improved.

Key words: GPS, uncertainty, standard-chain

1 Introduction

Since the establishment in 1996, ISO/TC213 has been working to develop the geometrical product specification (GPS) system, which links the whole course of geometrical products from specification, manufacturing to verification [1]. The "uncertainty" is used as an economic tool to enable optimum allocation of resource amongst specification, manufacturing and verification. In the GPS system, correlation uncertainty, specification uncertainty, compliance uncertainty and total uncertainty are defined besides measurement uncertainty [2]. The relationship of various uncertainties is shown in fig. 1.

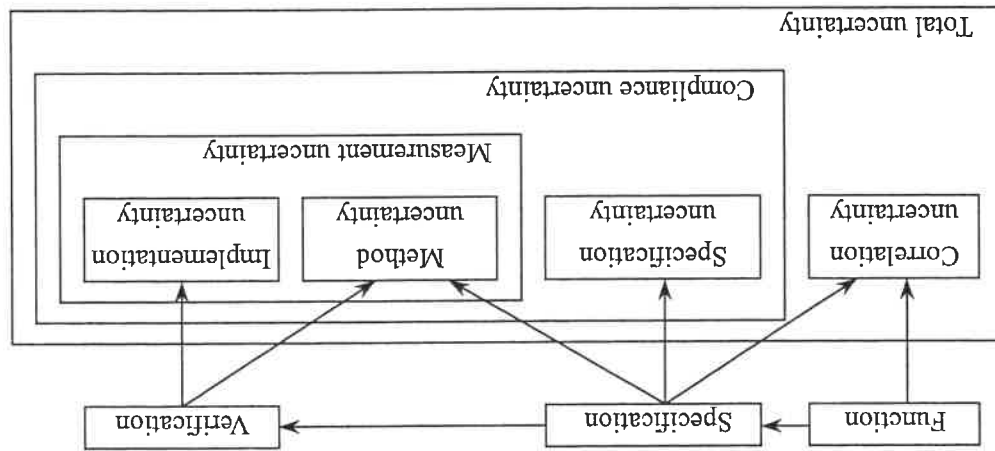


Fig. 1 Relationship of various uncertainties

2 Modeling of GPS standard-chain

Correlation uncertainty is usually not related to one single GPS specification. Commonly it takes a number of single GPS specifications to simulate a function. Therefore, for a given single GPS specification, the most important question is to determine compliance uncertainty of the corresponding GPS standard-chain, which is used to prove conformance or non-conformance of feature characteristic. Compliance uncertainty is the sum of specification uncertainty and measurement uncertainty, with which it can be proven that a workpiece complies with all possible interpretation of a specification. The computation method of measurement uncertainty is given in GUM [3] and ISO 14253-2 [4]. However, there is no specific method to calculate the compliance uncertainty until now. In order to solve this urgent question, a framework for the compliance uncertainty evaluation of GPS standard-chain is presented in this article.

A GPS standard-chain consists a set of standards related to a given GPS specification. The standards are collected in a few groups – “links” of the chain. There are 7 chain links associated with different geometrical characteristics of features [5]. It is clear that each GPS standard-chain is related to the complete process of specification, manufacturing and verification. Specification is the primary variation-control tool in design and verification is the primary tool to assess conformance to specification. A manufacturing process is selected to realize the design in a cost effective manner and to check the geometrical characteristic accuracy. In a GPS standard-chain, geometrical feature occurs in three disciplines: nominal feature, specification feature and verification feature, which correspond to nominal surface model, skin model and sampled surface model respectively.

2.1 Nominal surface model

Nominal surface model is the surface model of ideal geometry defined by the technical product documentation. The integral feature part of a nominal surface model is nominal feature NF , which is an ideal geometric entity.

2.2 Skin model

Skin model is the surface model of non-perfect geometry. The integral feature part of a skin model is specification feature SF , which is a non-ideal geometric entity. The skin model and SF are accordant with specification operator, which is an ordered set of such specification operations as partition, extraction, filtration, association, collection, construction and evaluation. The specification operator can be characterized by a function F_{sp} , so the input of F_{sp} is SF and the output is the specification of characteristic T_{sp} . Accordingly, every operation of the specification operator can be looked on as a sub-function of F_{sp} .

2.3 Sampled surface model

Sampled surface model is approximation of the discrete surface model obtained by sampling of the workpiece with measuring instruments. The integral feature part of a sampled surface model is verification feature VF , which is a non-ideal geometric entity. The sampled surface model and VF are accordant with verification operator, which is an ordered set of such verification operations as partition, extraction, filtration, association, collection, construction and evaluation. The verification operator can be characterized by a function F_{vp} , so the input of F_{vp} is VF and the output is the evaluation of characteristic S_{vp} . Accordingly, every operation of the verification operator can be looked on as a sub-function of F_{vp} .

2.4 Comparison for conformance

The specification operator is a default operator, and the verification operator is accordant with the specification operator completely. In this state, the method uncertainty equals to zero, so the compliance uncertainty should be calculated by the propagation of the implementation uncertainty through the specification

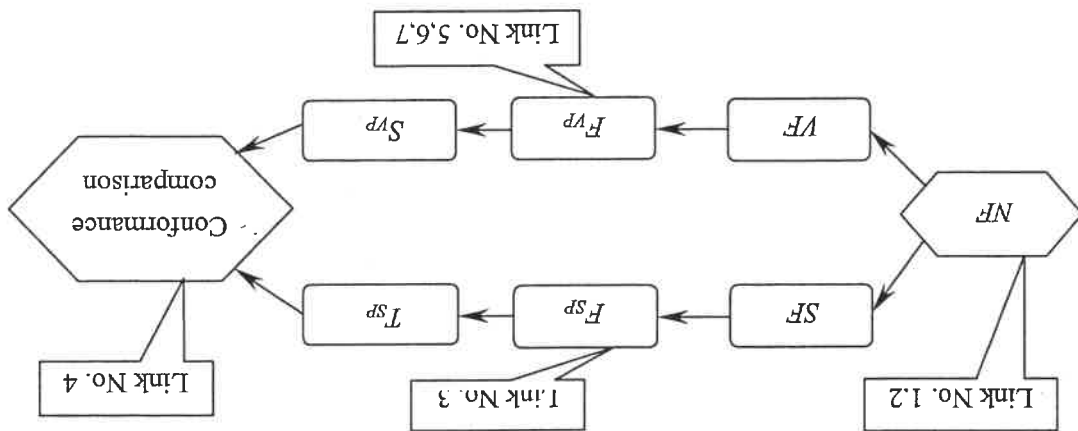
3.1 Default state

compliance uncertainty is proposed respectively as follows.

operator is accordant with the verification operator. Aiming at the two states, the flow for the computation of in default state or special state. The biggest difference between the two states is that whether the specification uncertainty is not existent until now. According to ISO 17450-2, a GPS process should be either was ignored. Although the GPS system gives the definition of specification uncertainty, the ISO standard for standard-chain. In the past, the measurement uncertainty was regarded mainly but the specification uncertainty and measurement uncertainty arise from many causes and propagate through the GPS standard-chain. The determination of compliance uncertainty is a complex question because specification uncertainty is to calculate the compliance uncertainty for a given GPS

3 Compliance uncertainty of GPS standard-chain

Fig. 2 Model of GPS standard-chain



Summarizing above-mentioned, the model of GPS standard-chain can be shown as in figure 2.

standard-chain, which takes into account specification uncertainty as well as measurement uncertainty.

is not reduced drastically. Therefore, the decision rules should be based on compliance uncertainty for a GPS performed with small measurement uncertainty and of no use, if the specification uncertainty documentation, is much larger than the measurement uncertainty. In this instance, the measurements sometimes the specification uncertainty evaluated for specifications given on existing product uncertainty is not taken into account, which are existent actually in GPS specifications. Furthermore, for the supplier to accept the workpiece. Obviously, the decision rules are improper because the specification measured values are inside the "grey-zones", it is neither possible for the customer to reject the workpiece, nor prove conformance, and only measured values in the non-conformance zone can prove non-conformance. According to the decision rules given in ISO 14253-1^[6], only measured values in the conformance zone can The measurement uncertainty causes "grey-zones" around the specification limits given on a drawing.

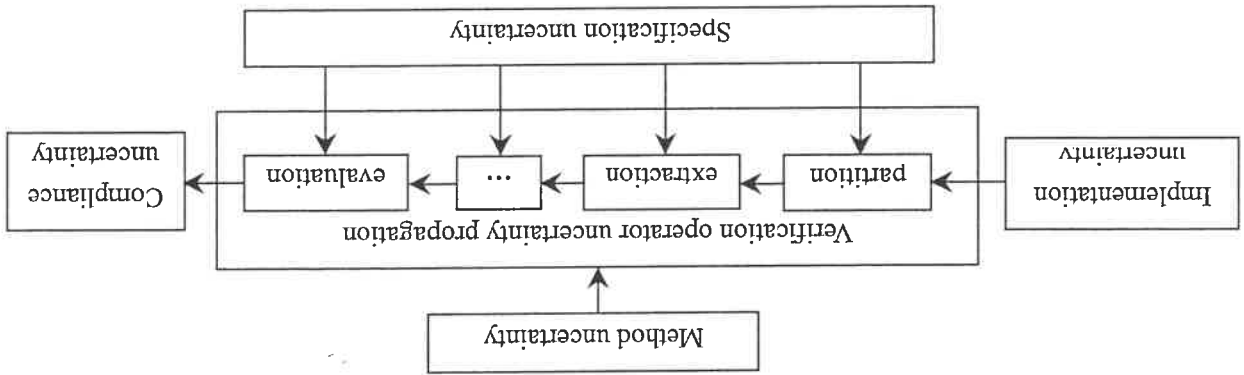
- partition from the skin model of a non-ideal surface;
 - partition of non-ideal lines from this non-ideal surface;
 - extraction using the evaluation length given in ISO 4288;
 - filtration using Gaussian filter with cut-off wavelength determined by the rules in ISO 4288 and the corresponding stylus tip radius and sample spacing shall be used;
 - evaluation of R_p values as defined in ISO 4287 and 4288.
- According to ISO 17450-2, the default specification operator of R_p 16.0 μ m indicates that:

4.1 Modeling of GPS standard-chain corresponding to R_p 16.0 μ m

In this section, a case study for the GPS specification of R_p 16.0 μ m is given in the default state.

4 A case study

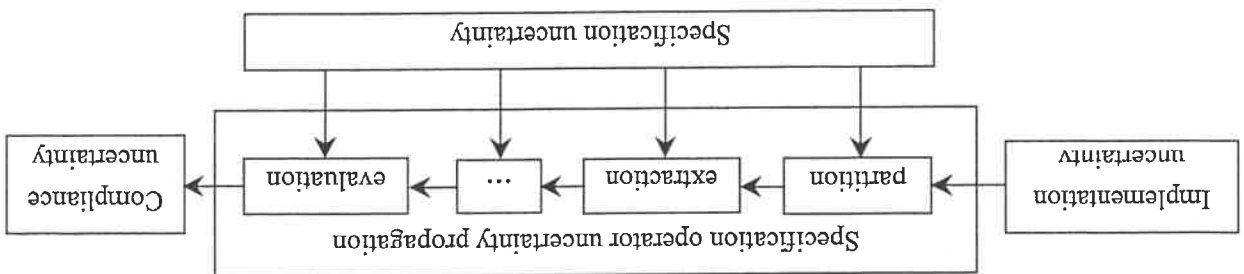
Fig. 4 Calculation flow for compliance uncertainty in special state



The specification operator is a special operator, and/or the verification operator is not accordant with the specification operator. In this state, the method uncertainty doesn't equal to zero commonly, so the compliance uncertainty should be obtained by the propagation of the implementation uncertainty through the verification operator, in which both the specification uncertainty and method uncertainty are inherent, as shown in fig. 4.

3.2 Special state

Fig. 3 Calculation flow for compliance uncertainty in default state



operator or verification operator, in which the specification uncertainty is inherent, as shown in fig. 3.

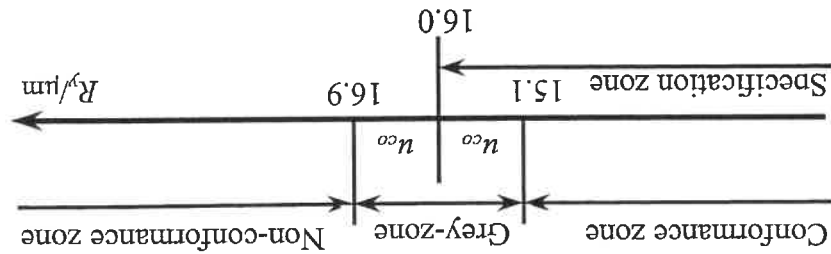


Fig. 5 Result of conformance comparison

conformable to the measurand.

In conformity to the decision rules given in 2.4, the result of conformance comparison is shown in fig. 5. It can be seen that the measured value ($R_y = 15.0\mu\text{m}$) is inside the conformance zone, so the measured value is

4.3 The result of conformance comparison

$$n_{co} = \sqrt{n_f^2 + n_{f,i}^2 - 2\rho_{f,i}n_f} = 0.9\mu\text{m} \quad (3)$$

by the uncertainty propagation formula

The last step: The filtered data are correlated, although the input data are totally uncorrected. The reason is that each filtered value, due to the convolution process, is calculated from several input values in the neighborhood of this particular value. The compliance uncertainty after evaluation operation can be calculated

$$n_f = \sqrt{1 - \frac{\Delta_x}{a\lambda_c} \left(2 - \frac{\sqrt{2}}{1} \right) n_{im}} = 0.646\mu\text{m} \quad (2)$$

operation can be expressed as

According to the uncertainty propagation rule of Gaussian filter [8], the uncertainty n_f after the filtration

$$s(x) = \frac{1}{\alpha\lambda_c} \exp \left[-\pi \left(\frac{x}{\alpha\lambda_c} \right)^2 \right] \quad (1)$$

weighting function $s(x)$ defined by the equation

The second step: The Gaussian filter according to ISO 11562 [7] is a linear profile filter of a continuous

which is equated to be $0.648\mu\text{m}$.

The first step: Usually, it is supposed that the sampling points are totally uncorrelated and the uncertainty of sampling points is equal, so the implementation uncertainty n_{im} is the uncertainty of single sampling point,

operation, which is the final result of compliance uncertainty.

According to the flow chart in default state, the calculation of compliance uncertainty can be divided into three steps: the first step, to determine the implementation uncertainty; the second step, to determine the uncertainty after the filtration operation and the last step, to determine the uncertainty after evaluation operation, which is the final result of compliance uncertainty.

4.2 Compliance uncertainty of the GPS standard-chain

data are filtered, the result of R_y can be numerated, which equals to $15.0\mu\text{m}$.

sampling length $l_s = 2.5\text{mm}$, evaluation length $l_n = 5\text{mm}$ and cut-off wavelength $\lambda_c = 2.5\text{mm}$. After sampling The specification of characteristic is measured by the stylus instrument with sampling interval $\Delta x = 5\mu\text{m}$, completely, which consists of partition, partition, extraction, filtration, evaluation orderly, too.

Because it is in the default state, the verification operator is accordant with the specification operator

5 Conclusions

A framework for uncertainty evaluation of GPS standard-chain is proposed in this article, which is based on the modeling of GPS standard-chain. It enables us to generate compliance uncertainty on the verification of GPS standard-chain, which makes the acceptance or refusal of feature characteristic can be conducted in a quantitative way, so the veracity of verification could be improved. Furthermore, it can be seen clearly that the calculation process of compliance uncertainty is easier in the default state than the special state because the specification operator is accordant with the verification operator. So the "default" should be used as much as possible in a GPS standard-chain, which embodies the simplification principle advocated by the GPS system.

Acknowledgements

The author would like to thank the Ministry of National Science and Technology for the support of this work under the grant 2002BA906A20.

References

- [1] ISO/TS 17450-1, 2002, Geometrical product specification (GPS) – General concept – Part 1: Model for geometric specification and verification.
- [2] ISO/TS 17450-2, 2002, Geometrical product specification (GPS) – General concept – Part 2: Basic tenets, specifications, operators and uncertainties.
- [3] Guide to the Expression of Uncertainty in Measurement (GUM). BIPM, IEC, IFCC, ISO, IUPAP, IUPAC, OIML, 1st edition, 1995.
- [4] ISO/TS 14253-2, 1999, Geometrical product specification (GPS) – Inspection by measurement of workpieces and measuring equipment – Part 2: Guide to the estimation of uncertainty in GPS measurement, in calibration of measuring equipment and in product verification.
- [5] ISO/TR 14638:1995, Geometrical product specification (GPS)—Masterplan.
- [6] ISO 14253-1, 1998, Geometrical product specification (GPS) – Inspection by measurement of workpieces and measuring equipment – Part 1: Decision rules for proving conformance or non-conformance with specifications.
- [7] ISO 11562, 1996, Geometrical Product Specifications (GPS) – Surface Texture: Profile Method – Metrological Characteristics of Phase Correct Filters.
- [8] Michael Krystek. Measurement uncertainty propagation in the case of filtering in roughness measurement. Measurement Science and Technology. 12 (2001) 63 – 67

Author Index

Arenberg, Jonathan W.	40	Mink, Ronald G.	72
Baldwin, Jon M.	62	Novak, Erik L.	83
Blake, Peter N.	72	Phillips, Steven D.	5, 14
Bönsch, Gerhard	46	Rhorer, Richard L.	106
Byun, Jeongmin	122	Robinson, F. David	72
Campbell, Daniel A.	62	Sato, Takashi	134
Canning, J. Suzanne	113	Schmitz, Tony L.	78, 113
Chambers, V. John	72	Schödel, René	46
Content, David	72	Schwenke, Heinrich I.	140
Davies, Angela D.	78	Sedransk, Nell	100
Davila, Pamela	72	Shah, Dilip A.	34
Decker, Jennifer E.	46	Shimjo, Makoto	134
Doiron, Theodore D.	28	Siebert, Bernd R. L.	56
Drescher, Joseph D.	107	Smith, Philip T.	128
Estler, W. Tyler	3, 14	Sommer, Klaus-Dieter	56
Evans, Christopher J.	89	Strawderman, William E.	100
Frank, Matthias	140	Summerhays, Kim D.	62
Frey, Eric	83	Tanaka, Hideyuki	134
Guthrie, Richard W.	95	Tolman, B.	14
Henke, Richard P.	62	Trenk, Michael	140
Holmes, Michael L.	89	Vallance, R. Ryan	128
Iyer, Han K.	68	van Eijk, Jan	50
Jiang, X. Q.	146	Wan, Der-Shen	83
Kochsiek, Manfred	56	Wang, C. M. Jack	68
Li, Zhu	146	Wang, Jin Xing	146
Liu, C. Richard	122	Weckenmann, Albert A.	56
Liu, Hung-kung	100	Xu, Z. G.	146
Lu, Z. Q. John	95	Zhang, Nien F.	100
Ma, L. M.	146	Ziegert, John C.	113
Meyer, Piotr J.	50		

